

**T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**UYDU GÖRÜNTÜLERİNİN İŞLENMESİNDE KULLANILAN
YARDIMCI VERİLER VE SINIFLANDIRMA DOĞRULUĞUNA
ETKİLERİ**

**FURKAN BİLÜCAN
YÜKSEK LİSANS TEZİ
HARİTA MÜHENDİSLİĞİ ANA BİLİM DALI
JEODEZİ VE COĞRAFİ BİLGİ TEKNOLOJİLERİ PROGRAMI**

**GEBZE
2021**

**T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**UYDU GÖRÜNTÜLERİNİN İŞLENMESİNDE
KULLANILAN YARDIMCI VERİLER VE
SINIFLANDIRMA DOĞRULUĞUNA ETKİLERİ**

**FURKAN BİLÜCAN
YÜKSEK LİSANS TEZİ
HARİTA MÜHENDİSLİĞİ ANA BİLİM DALI
JEODEZİ VE COĞRAFİ BİLGİ TEKNOLOJİLERİ PROGRAMI**

**DANIŞMANI
PROF. DR. TAŞKIN KAVZOĞLU**

**GEBZE
2021**

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**ANCILLARY AND AUXILIARY DATA USED IN
THE PROCESSING OF SATELLITE IMAGES AND
THEIR EFFECTS ON CLASSIFICATION ACCURACY**

FURKAN BİLÜCAN

**A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE DEPARTMENT OF
GEOMATICS ENGINEERING
GEODESY AND GEOGRAPHICAL INFORMATION TECHNOLOGIES PROGRAMME**

**THESIS SUPERVISOR
PROF. DR. TAŞKIN KAVZOĞLU**

GEBZE

2021

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 10/06/2021 tarih ve 2021/26 sayılı kararıyla oluşturulan jüri tarafından 24/06/2021 tarihinde tez savunma sınavı yapılan Furkan BİLÜCAN'ın tez çalışması Harita Mühendisliği Anabilim Dalı, Jeodezi ve Coğrafi Bilgi Teknolojileri Programında Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI)

: Prof. Dr. Taşkın KAVZOĞLU

ÜYE

: Prof. Dr. Fevzi KARSLI

ÜYE

: Doç. Dr. İsmail ÇÖLKESEN

ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
...../...../..... tarih ve/..... sayılı kararı.

İMZA/MÜHÜR

ÖZET

Gelişen uydu teknolojileri sayesinde uzaktan algılama teknikleriyle toplanan verinin sayısı, türü ve hacmi önemli ölçüde artmıştır. Bu veriler çevre kirliliğinin belirlenmesi, su kalitesi değişim analizinin yapılması, tarımsal alanların izlenilmesi, kaçak yapıların belirlenmesi ve arazi kullanım/arazi örtüsü haritalarının üretilmesi gibi birçok farklı uygulamada kullanılmaktadır. Tez çalışması kapsamında orta çözünürlüklü (10 m) ve ücretsiz olarak kullanıcılara sağlanan Sentinel-2A uydu görüntüsü kullanılarak yardımcı verilerin piksel tabanlı sınıflandırma doğruluğu üzerindeki etkileri araştırılmıştır. Bu amaç doğrultusunda yüksek doğrulukta arazi kullanım/arazi örtüsü haritaların üretilmesi için söz konusu uydu görüntüsüyle beraber birçok yardımcı veri değerlendirilmiştir. Bu yardımcı veriler uydu görüntüsünün spektral bantları kullanılarak üretilen doku özellikleri, morfolojik filtreler ile elde edilen özellikler, çeşitli spektral bantlar, temel bileşen analizi ile belirlenen bileşenler, ton-doygunluk-yoğunluk renk uzay dönüşümü ile elde edilen özelliklerdir. Dış kaynaklı yardımcı veriler ise küresel kentsel ayak izi haritası, nüfus haritası, eğim, bakı ve yükseklik haritaları, yeryüzüne ait toprak türü haritası, yağış miktarı ve toprak sıcaklığını temsil eden haritalardır. Yardımcı veriler ve uydu görüntüsünün beraber değerlendirilmesiyle 170 banda sahip yüksek boyutlu bir görüntü elde edilmiştir. Yüksek boyutlu görüntüler, tanımlamayı güçleştiren bantlar içerebildiğinden sınıflandırma işlemi aşamasında doğruluğu düşürmekte ve daha yüksek sayıda örnek sayısı ve işlem süresi gibi olumsuz etkileri beraberinde getirebilmektedir. Bu sorunu aşmak için filtre tabanlı, sarmal tabanlı ve gömülü özellik seçim yöntemleri uzaktan algılanmış veriye uygulanarak farklı boyutlarda alt kümeler belirlenmiştir. Belirlenen alt kümelerin, rastgele orman, destek vektör makineleri ve aşırı gradyan artırma makineleri (XGBoost) algoritmalarıyla ve bunların bir arada değerlendirildiği oylama tekniğiyle sınıflandırma analizleri yapılmıştır. Sınıflandırma sonuçları McNemar's doğruluk testi ile analiz edilerek istatistiksel olarak sonuçlar arasındaki farklılıklarının anlamlı olup olmadıklarının analizi yapılmıştır.

Anahtar Kelimeler: Uzaktan Algılama, Yardımcı Veri, Görüntü Sınıflandırma, Özellik Seçim, Makine Öğrenmesi.

SUMMARY

Thanks to recent developments in satellite technologies, number, type and volume of data provided by remote sensing techniques has increased significantly. These data are used in many applications, such as environmental pollution determination, water quality change analysis, monitoring of agricultural areas, determination of unlicensed construction and production of land use/land cover maps. As part of the dissertation, the effects of auxiliary data on pixel-based classification accuracy were investigated using a medium-resolution (10 m) Sentinel-2A satellite image provided to users for free. For this purpose, many auxiliary data were evaluated along with the satellite image in question to produce high-accuracy land use/land cover maps. These ancillary data are texture features, properties obtained by morphological filters, several spectral bands, components determined by principal component analysis, properties obtained by hue–saturation–intensity color space transformation. On the other hand, auxiliary data include global urban footprint map, population map, slope, aspect and elevation maps, soil type map of the Earth cover, precipitation, and soil temperature maps. A hyperspectral image consisting of 170 spectral bands was obtained by evaluating the ancillary and auxiliary data together with the satellite image. Hyperspectral images can cause problems during the classification process due to containing bands that make it difficult to identify. In order to overcome this problem, filter, wrapper and embedded feature selection methods were applied to obtain subsets. The subsets were classified by random forest (RF), support vector machines (SVM) and extreme gradient boosting machines (XGBoost) algorithms and then the results are combined by voting technique. The classification results were examined using McNemar's statistical test to validate whether differences results are significant or not.

Key Words: Remote Sensing, Ancillary and Auxiliary Data, Image Classification, Feature Selection, Machine Learning.

TEŞEKKÜR

Başta, değerli görüşleri ile tez konusunun belirlenmesinden, tez çalışmasının tamamlanmasına kadar geçen süreçte bana yol gösteren ve benden desteğini esirgemeyen kıymetli tez danışmanım Sayın Prof. Dr. Taşkın KAVZOĞLU'na,

Tez çalışmam boyunca kıymetli görüşleri ile bana yardımcı olan değerli hocam Sayın Doç Dr. İsmail ÇÖLKESEN'e,

Değerli çalışma arkadaşlarım ve tüm GTÜ Harita Mühendisliği Bölümündeki hocalarıma,

Benden desteklerini esirgemeyen aileme şükranlarımı ve saygılarımı sunuyorum.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	v
SUMMARY	vi
TEŞEKKÜR	vii
İÇİNDEKİLER	viii
SİMGELER ve KISALTMALAR DİZİNİ	xi
ŞEKİLLER DİZİNİ	xiii
TABLolar DİZİNİ	xv
1. GİRİŞ	1
1.1. Tezin Amacı ve İçeriği	3
1.2. Literatür Araştırması	6
2. UZAKTAN ALGILAMADA TEMEL KAVRAMLAR	11
2.1. Elektromanyetik Enerji ve Elektromanyetik Spektrum	11
2.2. Uzaktan Algılamada Çözünürlük Kavramı	13
2.3. Hiperspektral (Yüksek Boyutlu) Görüntüler	14
2.4. Görüntü Dönüşümleri	16
2.4.1. Doku Özellikleri	16
2.4.2. Temel Bileşenler Analizi	18
2.4.3. Ton-Doygunluk-Yoğunluk Renk Uzay Dönüşümü	20
2.4.4. Spektral İndisler	21
2.4.5. Morfolojik Filtreler	23
3. UZAKTAN ALGILAMADA SINIFLANDIRMA	25
3.1. Kontrollü ve Kontrolsüz Sınıflandırma	25
3.2. Sınıflandırma Algoritmaları	27
3.2.1. Maksimum Olabilirlik Yöntemi	29
3.2.2. Destek Vektör Makineleri	29
3.2.3. Rastgele Orman Algoritması	31
3.2.4. XGBoost Algoritması	32
3.2.5. Oylama Tekniği	33

3.3. Sınıflandırma Sonuçlarının Analizi	34
3.3.1. McNemar's Testi	37
3.4. Ters Mesafe Ağırlıklı Enterpolasyon Yöntemi (IDW)	37
4. ÖZELLİK SEÇİMİ	39
4.1. Filtre Tabanlı Yöntemler	40
4.1.1. Ki-kare Testi	41
4.1.2. ReliefF	41
4.1.3. Bilgi Kazancı	42
4.2. Sarmal Tabanlı Yaklaşımlar	43
4.2.1. Genetik Algoritma	45
4.2.2. Ardışık ileri Yönde Seçim Yaklaşımı	46
4.2.3. Özyinelemeli Özellik Seçim Yaklaşımı	47
4.3. Gömülü Yöntemler	47
4.3.1. LightGBM ile Özellik Seçimi	49
4.3.2. Hilbert-Schmidt Serbestlik Kriteri Lasso ile Özellik Seçimi	50
4.3.3. Rastgele Orman Algoritması ile Özellik Seçimi	51
5. ÇALIŞMA ALANI ve VERİ SETİ HAZIRLAMA	52
5.1. Spektral Bantlardan Üretilen Yardımcı Veriler	55
5.1.1. Gri Düzey Eş Oluşum Matrisleri ile Üretilen Doku Özellikleri	55
5.1.2. Temel Bileşenler Analizi	55
5.1.3. Ton-Doygunluk-Yoğunluk Renk Uzay Dönüşümü	57
5.1.4. Morfolojik Filtreler	58
5.1.5. Spektral İndisler	59
5.2. Diğer Yardımcı Veriler	67
5.2.1. Küresel Kentsel Ayak İzi	68
5.2.2. Nüfus	69
5.2.3. Eğim, Bakı ve Yükseklik	70
5.2.4. Toprak Sıcaklığı	72
5.2.5. Yağış Miktarı	73
5.2.6. Toprak Türü	74
6. UYGULAMA	75
6.1. Uydu Görüntüsünün Sınıflandırılması	76
6.2. Filtre Tabanlı Boyut İndirgeme	78

6.2.1. Ki-kare Yöntemi ile Boyut İndirgeme	79
6.2.2. ReliefF Yöntemi ile Boyut İndirgeme	82
6.2.3. Bilgi Kazancı Yöntemi ile Boyut İndirgeme	84
6.2.4. Sonuçların Karşılaştırılması	87
6.3. Sarmal Tabanlı Boyut İndirgeme	90
6.3.1. Genetik Algoritma	91
6.3.2. Ardışık İleri Yönde Özellik Seçim Yaklaşımı	93
6.3.3. Özyinelemeli Özellik Seçim Yaklaşımı	93
6.3.4. Sonuçların Karşılaştırılması	95
6.4. Gömülü Boyut İndirgeme	98
6.4.1. LightGBM ile Boyut İndirgeme	99
6.4.2. Hilbert-Schmidt Serbestlik Kriteri Lasso ile Boyut İndirgeme	100
6.4.3. Rastgele Orman Algoritması ile Boyut İndirgeme	102
6.4.4. Sonuçların Karşılaştırılması	105
6.5. McNemar's Testi ile Sınıflandırma Sonuçlarının Analizi	108
7. SONUÇLAR ve TARTIŞMA	110
KAYNAKLAR	116
ÖZGEÇMİŞ	128

SİMGELER ve KISALTMALAR DİZİNİ

<u>Simgeler ve</u> <u>Kısaltmalar</u>	<u>Açıklamalar</u>
χ^2	: Ki-kare
λ	: Dalga boyu
ν	: Frekans
AİS	: Ardışık ileri yönde seçim yaklaşımı
BAEI	: Yerleşim alanı çıkarım indisi
BRBA	: Yerleşim alanı için bant oranı
c	: Işık hızı
DBSI	: Kuru çıplak-toprak indisi
DVM	: Destek vektör makineleri
E	: Elektrik alanı
EOTA	: Ekili olmayan tarım alanı
ETA	: Ekili tarım alanı
EVI	: Zenginleştirilmiş bitki örtüsü indisi
GA	: Genetik algoritma
GARI	: Yeşil atmosferik dirençli bitki örtüsü indisi
CIg	: Yeşil klorofil indisi
GEMI	: Küresel çevre izleme indisi
GLI	: Yeşil yaprak indisi
GNDVI	: Yeşil normalize edilmiş fark bitki örtüsü indisi
GUF	: Küresel kentsel ayak izi
H	: Ton
HSIC Lasso	: Hilbert-Schmidt Serbestlik Kriteri Lasso
I	: Yoğunluk
IDW	: Ters mesafe ağırlıklı enterpolasyon yöntemi (IDW)
IFOV	: Anlık görüş açısı
InfG	: Bilgi Kazancı
IPVI	: Kızılötesi oran bitki örtüsü indisi
LightGBM	: Hafif gradyan artırma makineleri algoritması

<i>M</i>	: Manyetik alan
<i>m</i>	: metre
MIR	: Orta kızılötesi
MSI	: Nem stres indisi
NDTI	: Normalize edilmiş fark toprak sürme indisi
NDVI	: Normalize edilmiş fark bitki indisi
NDWI	: Normalize edilmiş fark su indisi
NIR	: Yakın kızılötesi
OSAVI	: Optimize edilmiş toprak ayarlı bitki örtüsü indisi
ÖÖS	: Özyinelemeli özellik seçim yaklaşımı
RDVI	: Yeniden normalize edilmiş bitki örtüsü indisi
RENDVI	: Kırmızı kenar normalize edilmiş fark bitki indisi
RO	: Rastgele orman algoritması
<i>S</i>	: Doygunluk
SAR	: Yapay Açıklıklı Radar
SATVI	: Toprak ayarlı toplam bitki örtüsü indisi
SAVI	: Toprak bitki örtüsü indisi
SIPI	: Yapı duyarlı pigment indisi
SYM	: Sayısal yükseklik modeli
TB	: Temel bileşen
TDVI	: Dönüştürülmüş fark bitki örtüsü indisi
VARI	: Görünür atmosferik direnç indisi
WRI	: Su oranı indisi
XGB	: Aşırı gradyan artırma makineleri algoritması
YSA	: Yapay sinir ağları

ŞEKİLLER DİZİNİ

<u>Sekil No:</u>	<u>Sayfa</u>
1.1: Tez çalışması kapsamında takip edilen iş akış şeması.	5
2.1: Elektromanyetik dalgalar.	11
2.2: Elektromanyetik spektrum.	12
2.3: Hiperspektral görüntü.	15
2.4: a) RGB renk uzayı, b) HSI renk uzayı.	20
3.1: Kontrollü ve kontrolsüz sınıflandırma.	26
3.2: Destek vektör makinelerinin iki sınıfı ayırmak için kullanımı.	30
3.3: Rastgele orman algoritması çalışma ilkesi.	32
3.4: Seviye odaklı ağaç büyümesi.	33
3.5: Oylama yöntemi.	34
3.6: Hata matrisi gösterimi.	35
4.1: Filtre tabanlı özellik seçim yöntemi.	40
4.2: Sarmal özellik seçim yaklaşımı.	44
4.3: Genetik algoritma yaklaşımı.	46
4.4: Gömülü özellik seçim yöntemi.	48
4.5: Yaprak odaklı ağaç büyümesi.	50
5.1: Çalışma alanı.	52
5.2: Temel bileşen analizi ile hesaplanan varyans değerleri.	56
5.3: NIR-R-G bant kombinasyonu ile hesaplanan doygunluk haritası.	57
5.4: a) Genişleme morfolojik filtresi, b) Aşınma morfolojik filtresi, c) Açılma morfolojik filtresi, d) Kapanma morfolojik filtresi.	59
5.5: Çalışma alanına ait NDVI haritası.	60
5.6: Çalışma alanına ait küresel kentsel ayak izi haritası.	69
5.7: Çalışma alanına ait nüfus haritası.	70
5.8: Çalışma alanına ait eğim haritası.	71
5.9: Çalışma alanına ait bakı haritası.	71
5.10: Çalışma alanına ait yükseklik haritası.	72
5.11: Verilerin temin edildiği istasyonlar.	73
5.12: Çalışma alanına ait toprak türü haritası.	74

6.1:	Sentinel -2A uydu görüntüsünün oylama tekniğiyle sınıflandırılması sonucu üretilen tematik harita.	77
6.2:	Sentinel -2A uydu görüntüsünün oylama tekniğiyle sınıflandırılması sonucu üretilen hata matrisi.	78
6.3:	Filtre tabanlı yöntemlerle belirlenen veri setlerine RO algoritmasının uygulanmasıyla hesaplanan genel doğruluk grafikleri.	88
6.4:	Ki-kare metoduyla belirlenen VS-7'nin oylama tekniğiyle üretilen tematik haritası.	89
6.5:	Ki-kare metoduyla belirlenen VS-7'nin oylama tekniğiyle sınıflandırılması sonucu üretilen hata matrisi.	90
6.6:	Sarmal yaklaşımlarla belirlenen veri setlerine RO algoritmasının uygulanmasıyla hesaplanan genel doğruluk grafikleri.	96
6.7:	GA ve RO uygunluk fonksiyonuyla belirlenen VS-5'in oylama tekniğiyle sınıflandırılması sonucu üretilen tematik haritası.	97
6.8:	GA ve RO uygunluk fonksiyonuyla belirlenen VS-5'in oylama tekniğiyle sınıflandırılması sonucu üretilen hata matrisi.	98
6.9:	Gömülü yöntemlerle belirlenen veri setlerine RO algoritmasının uygulanmasıyla hesaplanan genel doğruluk grafikleri.	106
6.10:	HSIC Lasso metoduyla belirlenen VS-4'in oylama tekniğiyle sınıflandırılması sonucu üretilen tematik haritası.	107
6.11:	HSIC Lasso metoduyla belirlenen VS-4'ün oylama tekniğiyle sınıflandırılması sonucu üretilen hata matrisi.	108

TABLÖLAR DİZİNİ

<u>Tablo No:</u>	<u>Sayfa</u>
2.1: Doku özellikleri ve formülleri.	18
5.1: Sentinel-2A uydusu teknik özellikleri.	53
5.2: Sentinel-2A uydusu spektral özellikleri.	54
6.1: Uydu görüntüsü sınıflandırma genel doğrulukları.	77
6.2: Ki-kare özellik seçim yöntemiyle belirlenmiş veri seti.	80
6.3: ReliefF özellik seçim yöntemiyle belirlenmiş veri seti.	83
6.4: InfG özellik seçim yöntemiyle belirlenmiş veri seti.	85
6.5: Filtre tabanlı yöntemlerle belirlenen veri setleri için hesaplanan sınıflandırma doğrulukları.	89
6.6: GA ve RO uygunluk fonksiyonuyla belirlenen küme.	92
6.7: ÖÖS ve RO uygunluk fonksiyonuyla belirlenen küme.	94
6.8: Sarmal yaklaşımlarla belirlenen veri setleri için hesaplanan sınıflandırma doğrulukları.	97
6.9: LightGBM özellik seçim yöntemiyle belirlenmiş veri seti.	100
6.10: HSIC Lasso özellik seçim yöntemiyle belirlenmiş veri seti.	101
6.11: RO özellik seçim yöntemiyle belirlenmiş veri seti.	103
6.12: Gömülü yöntemlerle belirlenen veri setleri için hesaplanan sınıflandırma doğrulukları.	106
6.13: Yapılan sınıflandırma işlemlerinin McNemar's istatistiksel testi ile analiz sonuçları.	109

1. GİRİŞ

Uzaktan algılama fiziksel temas olmaksızın nesneler hakkında bilgi sahibi olmamızı sağlayan ve teknolojinin giderek gelişmesi ile büyük gelişme gösteren önemli bir bilim dalıdır. Özellikle, yeryüzü ile ilgili çalışmalarda uydu görüntüleri kullanılarak çeşitli tematik haritalar üretilmektedir. Arazi örtüsü/arazi kullanımı haritaları üretmek için uygulanan sınıflandırma işleminde nasıl bir yol izleneceği ya da hangi verilerin kullanılması gerektiği soruları çalışmalar için büyük önem taşımaktadır. Bu durumun en önemli sebebi uzaktan algılama biliminin yeryüzüne ilişkin birçok çalışmada geniş bir kullanım yelpazesine sahip olmasıdır. Sınıflandırma işlemi sonucunda elde edilen tematik haritalara ait sınıflandırma doğrulukları, haritaların hangi oranda başarılı üretildiği ve ne kadar güvenilir oldukları hakkında bilgi vermektedir. Bunun doğal sonucu olarak kullanılan verinin nasıl elde edileceği ve elde edilen verinin işlendikten sonra nasıl analiz edileceği çalışmalar için büyük önem taşımaktadır. Uzaktan algılanmış görüntüler, arazi örtüsü/arazi kullanımı harita üretiminde ana veri kaynağı olarak kullanıldığı bilinmektedir [Salehi et al., 2012]. Bunun yanı sıra sınıflandırma doğruluğunu artıracak yardımcı verilerin kullanılması yüksek doğrulukta tematik harita elde etmek ve daha kullanışlı bir ürün ortaya çıkarmaya katkı sağlamaktadır [Rwanga and Ndambuki, 2017].

Literatürde, uydu görüntülerindeki piksellerin daha doğru sınıflandırılması için mevcut spektral bantlara ek olarak çeşitli yardımcı verilerin eklenmek suretiyle değerlendirildiği görülmektedir [Zhu et al., 2016], [Sluiter and Pebesma, 2010]. Uydu görüntüsünün spektral bantları kullanılarak çeşitli işlemler sonucunda elde edilen temel yardımcı veriler (ancillary data); gri düzey eş oluşum matris eşitliklerinin kullanımı ile hesaplanan doku özellikleri, temel bileşenler analizi ile tespit edilen bileşenler, ton-doygunluk-yoğunluk (HSI) renk uzayı dönüşümü ile elde edilen özellikler, spektral indisler ve morfolojik filtrelerin uygulanmasıyla elde edilen özelliklerdir. Bu özelliklerin yanında dış bir kaynaktan elde edilen çeşitli yardımcı veriler de (auxiliary data) literatürde kullanılmaktadır. Bu yardımcı veri grubuna; küresel kentsel ayak izi haritası, nüfus haritası, sayısal yükseklik modelinin değerlendirilmesiyle üretilen eğim, bakı ve yükseklik haritaları, yeryüzüne ait toprak türü haritası, yağış miktarı ve toprak sıcaklığını temsil eden haritalar örnek olarak verilebilir. Bununla beraber çok miktarda yardımcı verinin ve uydu görüntüsünün bir

arada değerlendirilmesiyle yüksek boyutta veri seti elde edilmektedir. Söz konusu yüksek boyutlu görüntüler içerisinde sınıflandırma işlemi için yarar sağlamayan veya aralarında yüksek korelasyon bulunan bantlar içerdiklerinden sınıflandırma işlemi aşamasında problem teşkil etmektedir. Yüksek boyutlu veriler daha fazla sayıda spektral bant içerdikleri için sınıflar arasındaki ayrılabilirliği mümkün kılmakla birlikte, sınıflandırma işleminde kullanılan çok fazla spektral bant sınıflandırma doğruluğunu azaltabilmektedir [Landgrebe, 2002]. Yüksek boyutlu verilerde karşılaşılan bu duruma boyutsallık sorunu (Hughes fenomeni, curse of dimensionality) adı verilmektedir [Hughes, 1968], [Belgiu and Drăgut, 2016]. Büyük veriler ile yapılan işlemlerdeki bir başka dezavantaj ise zaman maliyetidir ve bu sorunlara çözüm bulabilmek için özellik seçimi algoritmaları kullanılmaktadır. Yüksek boyutlu görüntülerde özellik seçimi sınıflandırma işlemi için uygun bantların belirlenmesi olarak ifade edilebilir [Jia et al, 2012].

Özellik seçimi, yüksek boyuta sahip olan söz konusu verilere daha düşük boyutta bir veri seti oluşturmak, işlem yükünü azaltmak ve daha yüksek sınıflandırma doğruluğu elde etmek amacıyla uygulanır [Singh et al., 2016]. Diğer bir ifadeyle özellik seçim tekniklerinin üç ana avantajının olduğu söylenebilir: Bunlardan ilki; sınıflandırma performansının artırılması, ikincisi; veri seti sayısının azaltılması ile yapılacak olan işlem için gereken zaman maliyetinin azaltılması ve son olarak; sınırlı sayıda eğitim verisinin mevcut olduğu durumlarda daha düşük veri seti kullanımına olanak sağlamasıdır [Kavzoglu, 2001]. Söz konusu özellik seçim yöntemlerinin incelenerek en etkili yöntemin belirlenmesi, uygun özellik sayısının tespit edilmesi ve tespit edilen özelliklere göre sınıflandırma yöntemlerinin performanslarının analiz edilmesi literatürde sıklıkla çalışılan önemli araştırma konularıdır. Özellik seçim teknikleri genel olarak üç ana başlıkta incelenebilir. Bunlar istatistiksel tabanlı filtreleme (filter), çeşitli sınıflandırma algoritmalarını kullanarak en iyi tahmin sonucuna göre yapılan sarmalama tabanlı (wrapper) ve sınıflandırma ile özellik seçimini yapısında barındırarak en uygun alt kümeyi seçmeye yarayan gömülü (embedded) yöntemlerdir [Saeys et al., 2007].

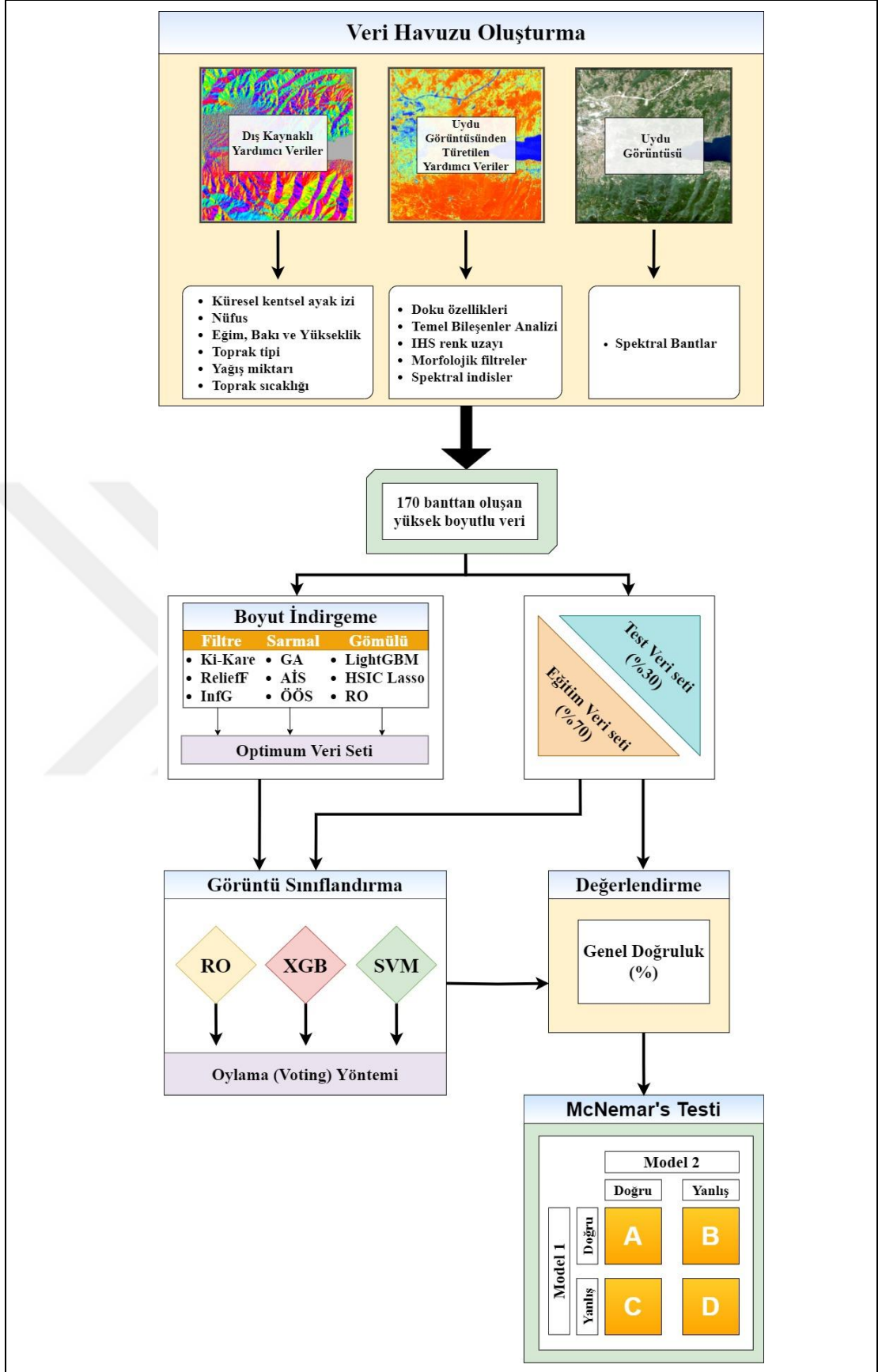
Yüksek doğrulukta tematik harita elde etme işleminde çeşitli yardımcı verilerin ve farklı özellik seçimi yöntemlerinin kullanılmasının yanında tercih edilen sınıflandırma algoritmaları da haritaların doğruluğu hakkında belirleyici bir rol oynar. Günümüze kadar uydu görüntülerinin işlenmesinde birçok sınıflandırma tekniği

görüntülere uygulanmış ve çeşitli başarı oranları elde edilmiştir. Bu sınıflandırma tekniklerinden istatistiksel tabanlı olarak çalışanlar (en çok benzerlik, paralel kenar vb.) istatistiksel olarak bazı kurallara göre uygulandığı için belirli bir dağılım göstermeyen verileri sınıflandırma hususunda yeterli başarıyı gösterememekte ve düşük doğrulukta sonuçlar üretmektedir. İstatistiksel tekniklerin yanında sınıflandırma işlemlerine uygulanmak üzere, Yapay Sinir Ağları gibi birçok sınıflandırma yöntemi geliştirilmiştir. Yapay Sinir Ağları istatistiksel tabanlı sınıflandırıcılara göre sınıfları tahmin etmede daha başarılıdır. Bununla beraber, son dönemlerde ise yapılan çalışmalarda makine öğrenme algoritmaları (destek vektör makineleri, rastgele orman, aşırı gradyan artırma vb.) uydu görüntüsü sınıflandırma işlemlerine başarıyla uygulanmaktadır. Yapay zeka alanının alt dallarından biri olan makine öğrenmesi algoritmaları veriler üzerinden öğrenip, tahminler yapıp model oluşturarak karışık sınıfları ayırt etmemize olanak sağlar. Bu özellikler sayesinde makine öğrenme algoritmaları birçok alanda olduğu gibi uzaktan algılama çalışmalarında da yoğun bir şekilde kullanılmaktadır. Uzaktan algılamada, özellikle yüksek boyuttaki veriler için makine öğrenme algoritmaları kullanılması geleneksel yöntemlere göre daha etkili performans göstermektedir [Huang et al., 2002], [Pal, 2005], [Ghimire et al., 2012].

1.1. Tezin Amacı ve İçeriği

Tez çalışması kapsamında, orta çözünürlüklü Sentinel-2A uydu görüntüsü kullanılarak yardımcı verilerin piksel tabanlı sınıflandırma doğruluğu üzerindeki etkileri incelenip detaylı bir analiz gerçekleştirilmesi amaçlanmıştır. Şekil 1.1’de tez çalışması için izlenilen iş akışında gösterildiği gibi çeşitli özellik seçim yöntemleriyle belirlenmiş olan bant kombinasyonlarının makine öğrenme algoritmaları (rastgele orman, aşırı gradyan artırma makineleri, destek vektör makineleri ve oylama metodu) ile sınıflandırılması ve sonuç olarak üretilen tematik haritaların doğruluklarının analiz edilmesi amaçlanmıştır. Orta çözünürlüklü Sentinel-2A uydu verisinin tercih edilmesinin temel nedenleri; ücretsiz olarak kullanıcılara sağlanması ve yeterli mekânsal çözünürlüğe (10 m) sahip olmasıdır. Yardımcı veriler ile görüntüyü zenginleştirip yüksek boyutlu bir görüntü elde etmek için mevcut bantlara ek olarak tüm yönlerde (0°, 45°, 90° ve 135°) her bant için zıtlık, korelasyon, farklılık, entropi, homojenlik, ortalama, açısal ikinci moment ve varyans doku özellikleri

hesaplanmıştır. Daha fazla yardımcı veri elde etmek için mevcut bantlara morfolojik filtrelerin uygulanmasıyla aşınma, genişleme, açılma ve kapanma işlemleri ile elde edilen özellikler de tez çalışması kapsamında değerlendirilmiştir. Bunun yanında bitki, su, toprak ve yerleşim alanlarının yapılarını ön plana çıkarmaya yarayan spektral bantlar, temel bileşenler ile belirlenen bileşenler, doygunluk-ton-yoğunluk renk uzayı dönüşümü ile elde edilen özellikler kullanılmıştır. Söz konusu bu yardımcı veriler mevcut bantlardan elde edilen yardımcı verilerdir. Bu verilere ek olarak bağımsız bir veri kaynağından edinilen çeşitli veriler de tez çalışmasında kullanılmıştır. Dış kaynaktan edinilen veriler; küresel kentsel ayak izi haritası, nüfus haritası, ALOS-PALSAR uydusundan türetilen eğim, bakı ve yükseklik haritaları, yeryüzüne ait toprak türü haritası, yağış miktarı ve toprak sıcaklığını temsil eden haritalardır. Ayrıca tez kapsamında kullanılan yardımcı veriler filtreleme (filter), sarmalama tabanlı (wrapper) ve gömülü (embedded) özellik seçimi yöntemleriyle en uygun bant kombinasyonun bulunabilmesi için birçok farklı veri seti elde edilmiştir. Değerlendirmeye alınan veri setleri için farklı makine öğrenme algoritmaları kullanılarak sınıflandırma işlemleri gerçekleştirilmiş ve sonuçlar McNemar's doğruluk testi ile analiz edilerek istatistiksel olarak performans farklılıklarının anlamlı olup olmadıklarının değerlendirilmesi yapılmıştır.



Şekil 1.1: Tez çalışması kapsamında takip edilen iş akış şeması.

1.2. Literatür Araştırması

Son yıllarda gelişen teknolojilerle beraber uzaktan algılama alanına olan ilgi hızlı bir artış göstermektedir. Uzaktan algılama bilimi, dünya yüzeyi, deniz veya okyanuslar hakkında yeri doldurulamaz bilgiler sağlarken geniş alanlar ile ilgili yapılan çalışmalarda uygun maliyet avantajı sağlamaktadır. Özellikle arazi örtüsü/arazi kullanımı haritaları üretmek için bu bilimin sahip olduğu avantajlardan yararlanılır. Örnek olarak uzaktan algılama verileri karbon emisyon gözlemlenmesinde [Birdsey et al., 2013], ormancılık ile ilgili çalışmalarda [Gong et al., 2013], kıyı çizgilerinde meydana gelen değişimlerin izlenmesinde [El-Asmar and Hereher, 2011], su kaynaklarının yönetilmesi ve su kalitesinin belirlenmesi gibi hidrolojik çalışmalarda [Topp et al., 2020], arazi değişim tespitinde [Grekousis et al., 2015], doğal afet analizinde [Khatami and Mountrakis, 2012], jeolojik harita üretiminde [Pour et al., 2018], kentsel alanların gelişiminde [Maktav et al., 2005] ve iklim değişikliği ile ilgili yapılan çalışmalarda kullanılmaktadır [Mortoja and Yigitcanlar, 2020]. Söz konusu çalışmaların gerçekçi bir bakış açısı ortaya koyabilmesi için altlık olarak kullanılan tematik haritaların üretilme doğruluğu ve doğruluk analizi oldukça önemlidir. Üretilen tematik haritaların doğruluğunu artırmak için birçok yardımcı verinin kullanıldığı görülmüştür. Literatürde uzaktan algılanmış görüntülerin sınıflandırılmasında doğruluğu artıracak çeşitli yardımcı veriler kullanıldığı görülmektedir.

Saha et al. [2005] tarafından yapılan bir çalışmada Normalize Edilmiş Fark Bitki İndisi (NDVI) ile sayısal yükseklik modelini (SYM) yardımcı veri olarak kullanmıştır. Uydu görüntüsünün mevcut bantlarına söz konusu yardımcı verilerin eklenmesiyle orijinal sınıflandırma doğruluğunun yaklaşık %4 oranda artış gösterdiği sonucuna ulaşılmıştır.

Diğer bir çalışmada Landsat-8 OLI uydu görüntüleri, NDVI, Düzeltilmiş Toprak Bitki İndisi (SAVI), SYM ile elde edilen plan eğriliği ile eğimin yanı sıra jeoloji ve iklim verileri kullanılmıştır. Kullanılan özellik seçimi yöntemlerinden birisi de özyinelemeli özellik seçim yöntemidir ve Rastgele Orman (RO) algoritması ile beraber kullanılmıştır. Sonuçta spektral değişkenler olarak adlandırılan bantlar ve indislere plan eğriliği ve eğim gibi yardımcı verilerin eklenmesi ile sınıflandırma doğruluğunun %3 artışı belirtilmiştir [Zeferino et al., 2020].

Bir başka çalışmada yardımcı veri olarak gri düzey eş oluşturma matris eşitlikleri kullanılarak hesaplanan doku özellikleri, vejetasyon indisleri ve ana bileşenler kullanılmıştır. Çeşitli veri setleri üretilerek sınıflandırıcı olarak destek vektör makineleri (DVM) algoritması seçilmiştir. Özellikle doku özelliklerinin sınıflandırma doğruluğuna etki ettiği belirtilmiş ve yaklaşık olarak %2,5 oranında doğruluğun iyileştiği ifade edilmiştir [Çölkesen ve Yomralıoğlu, 2014].

Amerika Birleşik Devletleri'nde Kaliforniya Eyaleti'nin San Diego şehrinde arazi örtüsü değişim analizinin yapıldığı bir çalışmada, 1990-1996 yılları arasına ait çok zamanlı Landsat TM uydu görüntülerine yükseklik, eğim ve yangın geçmişi gibi yardımcı verilerin birlikte değerlendirildiklerinde sınıflandırma genel doğruluğuna %15 oranında olumlu bir etki ettiği sonucuna ulaşılmıştır [Rogan et al., 2003].

Jin et al. [2018] tarafından orman ve çayır alanlarının karışık olduğu bir bölgede çok zamanlı uydu görüntüleriyle yapılan çalışmada, piksellerin daha doğru sınıflandırılması için doku özellikleri, Tasselled Cap dönüşümü ile elde edilen parlaklık, yeşillik ve ıslaklık eksenleri yardımcı veri olarak değerlendirilmiştir. Ayrıca SYM kullanılarak üretilen yükseklik, eğim ve bakı, Normalize Edilmiş Fark Bitki İndisi (NDVI), Modifiye Edilmiş Normalize Fark Su İndisi (MNDWI) yardımcı veri olarak kullanılmıştır. Dört farklı veri kümesi oluşturularak RO gömülü özellik seçim yöntemi sınıflandırma doğruluğuna en çok katkıda bulunan özellikleri belirlemek için kullanılmıştır. Bu veri kümelerinden çok zamanlı görüntülerin, eğimin, yüksekliğin, doku özelliklerinin ve NDVI bantlarının yardımcı veri olarak eklenmesiyle birlikte genel sınıflandırma doğruluğu %88.9 olarak hesaplanmıştır.

Sulak alanların nesne tabanlı sınıflandırılması amacıyla yapılan çalışmada yüksek çözünürlüklü insansız hava aracı Sentinel-2A verisi ile kaynaştırılarak sınıflandırma işlemi için yüksek performans elde edilmesi amaçlanmıştır. Bu bağlamda bant oranları ile hesaplanan çeşitli indisler, temel bileşenler, doku özellikleri, renk uzayı dönüşümünden elde edilen ton, doygunluk ve yoğunluk yardımcı veriler olarak değerlendirilmiştir. Sınıflandırma için genel doğruluğa en çok katkıda bulunan verilerin seçilmesi gömülü özellik seçimi RO algoritması ile gerçekleştirilmiş ve oluşturulan veri setinde özellik seçimi ile belirlenen 54 bant işleme alınarak, genel doğruluk %94.31 olarak hesaplanmıştır [Geng et al., 2020].

Hindistan'da bulunan Arjuni Havzası için yıllara göre kurak alan tespiti için yapılmış olan çalışmada, uzaktan algılama ile coğrafi bilgi sistemleri birlikte

kullanılmış ve yardımcı veri olarak varyans doku özelliği, Normalize Edilmiş Su Fark İndisi (NDWI) ile çalışma alanına ait drenaj ağından yararlanılmıştır. Çalışmada yıllara göre değerlendirilen görüntülerde yaklaşık olarak %5 oranında genel doğrulukta artış olduğu, söz konusu yardımcı verilerin sınıflandırma doğruluğuna olan olumlu etkileri ortaya konulmuştur [Thakkar et al., 2017].

Uydu görüntüsünün mevcut spektral bantlarıyla beraber renk uzay dönüşümleri, temel bileşenler, çeşitli vejetasyon indisleri, tüm yönlerde hesaplanan doku özelliklerinin yardımcı veri olarak kullanıldığı bir çalışmada, 96 bant içeren bir yüksek boyutlu bir veri seti oluşturulmuştur. Sınıflandırma sonucuna en çok katkı sağlayacak bantların belirlenmesi amacıyla ÖÖS yöntemini temel alan DVM algoritmasının kullanılmasıyla oluşturulan veri setleri ile en yüksek genel doğruluk %88,23 olarak hesaplanmıştır [Çölkesen ve diğ., 2015].

Tarımsal alanlarda ürün deseninin belirlenmesi ile ilgili nesne tabanlı sınıflandırma yönteminin kullanıldığı çalışmada yardımcı veri olarak tüm yönlerde hesaplanmış doku özelliklerinin yanı sıra NDVI kullanılmış ve %87,5 genel doğruluğun elde edildiği belirtilmiştir [Tavus vd., 2019].

Yapılan bir çalışmada yüksek çözünürlüklü Spot-5 uydu görüntüsü kural tabanlı sınıflandırma işlemi ile değerlendirilerek kentsel arazi örtüsü haritalarının daha yüksek bir doğrulukta hesaplanması amaçlanmıştır. Yardımcı veri olarak yol haritaları, kadastral haritalar, Openstreet haritaları, NDWI ve NDVI indisleri kullanılmıştır. Önerilen kural tabanlı sınıflandırma yöntemi ve kullanılan yardımcı verilerle kentsel alanlar için üretilen tematik haritaların %94 veya daha yüksek genel doğruluk ile elde edilebileceği belirtilmiştir [Sertel and Akay, 2015].

Spot-5 uydu görüntüsünün işlendiği, bant sayısının arazi örtüsü harita üretimindeki etkisinin araştırıldığı bir diğer çalışmada yardımcı veri olarak doku özellikleri, spektral ve bağlamsal özellikler ile vejetasyon indisleri kullanılmıştır. Yardımcı veriler ile oluşturulan görüntüye sınıflandırma ve regresyon karar ağaçları gömülü özellik seçimi uygulanarak üç farklı sınıflandırma algoritmasının performansları karşılaştırılmıştır ve destek vektör makinelerini sınıflandırma yöntemi olarak seçilerek %90,5 genel doğruluğa ulaşılmıştır [Myburgh and Niekerk, 2014].

Pleidas uydu görüntüleri ile Güney Afrika'nın Gauteng eyaletinde bulunan Boksburg ve Sunninghill şehirleri için yapılan bir çalışmada heterojen alanlar için arazi örtüsü sınıflandırılması işleminde matematiksel morfolojik tekniklerin önemi

araştırılmıştır. Morfolojik profillerin birleştirilmesiyle Boksburg şehri için üretilen haritanın genel doğruluğu %83,14, Sunninghill şehri için üretilen haritanın genel doğruluğu ise %83,19 olarak hesaplanmıştır. Multi-morfolojik profillerin uygulanmasıyla Boksburg şehri için %84,42 ve Sunninghill şehri için %84,08 oranında genel doğruluk hesaplandığı belirtilmiştir. [Tsoeleng et al., 2020].

Las Vegas ve Roma şehirlerine ait Quickbird uydu görüntülerinin ele alındığı bu çalışmada arazi kullanım haritaları elde etmek için morfolojik filtrelerin kullanılması ile görüntülerdeki yapılar hakkında gerekli bilgilerin sağlanması amaçlanmıştır. Doğruluğa katkıda bulunacak önemli bantların belirlenmesi amacıyla boyut azaltmak için ÖÖS yöntemini temel alan DVM algoritması kullanılmıştır. Seçilen bantlar esas alındığında morfolojik filtrelerin, iki görüntü için de sınıflandırma doğruluğu sonucuna pozitif etki ettiği sonucuna ulaşılmıştır [Tuia et al., 2009].

Yardımcı verilerin üretilip mevcut uydu görüntüsüne eklenmesi sonucu yüksek boyutlu görüntüler elde edildiği görülmektedir. Yüksek boyutlu bu görüntülerin ortaya çıkmasıyla uygun bir özellik seçim yaklaşımının belirlenmesi, özellik fazlalığını etkili bir şekilde azaltmak ve sınıflandırma verimliliğini iyileştirmek önemli bir araştırma konusudur. Bu yüzden literatürde yüksek boyutlu görüntülere uygulanan birçok özellik seçimi algoritması mevcuttur. Örnek çalışmalar aşağıda detaylı şekilde incelenmiştir.

Ki-kare, Bilgi Kazancı ve ReliefF filtre tabanlı özellik seçim yöntemleri uygun bantların belirlenmesi için uygulanmış ve yardımcı veriler ile üretilen yüksek boyutlu uydu görüntüsü dikkate alınmıştır. RO ve DVM'in sınıflandırıcı olarak kullanıldığı bu çalışmada Bilgi Kazancı özellik seçimi ile 110 banttan elde edilen 24 bant DVM algoritması ile sınıflandırıldığında %92,00 genel doğruluk elde edildiği ifade edilmiştir [Colkesen and Kavzoglu, 2018].

Geliştirilen yöntemlerle DVM algoritmasının, gömülü özellik seçimi olarak veya ÖÖS tekniğiyle birlikte verilere uygulanarak boyut indirgeme çalışmalarında değerlendirildiği bilinmektedir [Archibald and Fann, 2007], [Zhang and Ma, 2009].

Pal [2006] çalışmasında yüksek boyutlu veri setinden önemli bantların belirlenmesi için sarmal tabanlı özellik seçim yöntemi olan genetik algoritma (GA) ile uygunluk fonksiyonu olarak DVM sınıflandırma yöntemini, RO gömülü özellik seçim yöntemini ve DVM ile ÖÖS özellik seçim yöntemini kullanılmıştır. DVM/GA ile tespit edilen 15 ve 20 bantın sınıflandırma performansı diğer özellik seçim yöntemleri

ile elde edilen sonuçlarla karşılaştırılmıştır. DVM/GA özellik seçimi ile belirlenmiş 15 banttan oluşan veri seti için en yüksek sınıflandırma performansı %92,44 olarak hesaplanmıştır. Ayrıca 15 ve 20 banttan oluşan veri setlerinin sonuçları karşılaştırılarak, aralarında anlamlı bir fark olmadığı sonucuna ulaşılmıştır. DVM/GA özellik seçimi için eğitim aşaması farklı uygunluk fonksiyonları değerlendirildiğinde en yüksek olarak 602 dakikalık bir zaman maliyeti ortaya çıkmıştır. Ayrıca RO ile özellik seçim işlemi 12,98 saniye ve DVM-ÖÖS özellik seçim işlemi 3 dakikalık bir zaman maliyeti gerektirmiştir. Elde edilen bir diğer önemli sonuç GA için farklı uygunluk fonksiyonları söz konusu olduğunda hesaplama maliyetinin değiştiğidir.

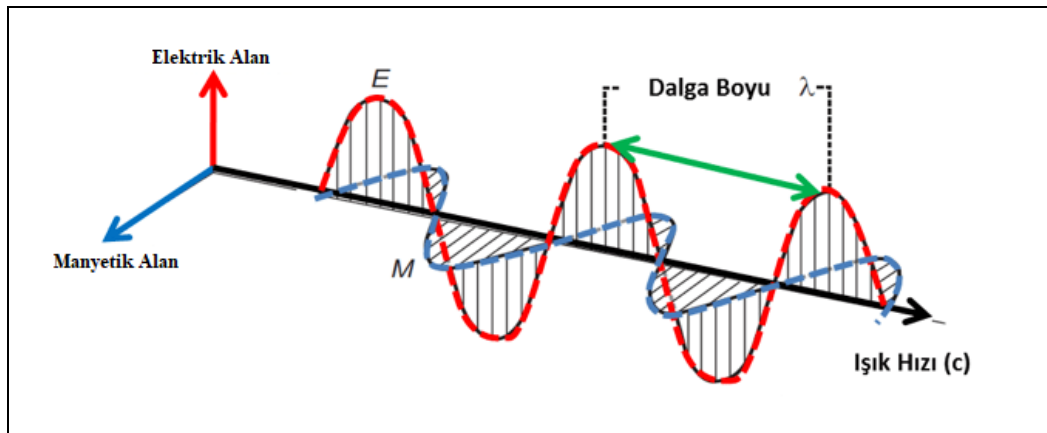
Bitkilerde hastalık tespit etmek amacıyla yapılan bir çalışmada ise en uygun bantların seçilmesi için DVM sınıflandırıcı ile genetik algoritma özellik seçim yöntemi kullanılmıştır. Yüksek boyutlu görüntüden seçilen 6 bant ile yapılan çalışmada sınıflandırma doğruluğu %97 olarak hesaplanırken; kırmızı, yeşil ve mavi görünür bantlar ile yapılan sınıflandırmada %76,92 oranında bir genel doğruluğa ulaşıldığı sonucuna ulaşılmıştır [Nagasubramanian et al., 2018].

Bitki fenotipinin belirlenmesi için yapılan bir çalışmada 215 banttan oluşan bir hiperspektral görüntüden optimum bantların seçilmesi amaçlanmıştır. Toplu özellik seçimi yöntemi üzerinde durulan bu çalışmada farklı özellik seçimi yöntemleri ele alınmıştır. Filtre tabanlı olan ReliefF ve korelasyon tabanlı özellik seçimi yöntemleri, sarmalama tabanlı olan öz yinelemeli özellik seçimi ile barebar DVM algoritması (DVM-ÖÖS), ardışık ileri yönde seçim yöntemi (AİS) son olarak gömülü özellik seçimi yöntemleri olan Lasso Regresyon ile RO algoritmaları analiz edilmiştir. En iyi sonuçların ReliefF, DVM-ÖÖS ve RO algoritmalarının kullanılmasıyla elde edildiği belirlenmiştir. Genel doğruluktaki iyileşme ise %8,5 olarak hesaplanmıştır [Moghimi et al., 2018].

2. UZAKTAN ALGILAMADA TEMEL KAVRAMLAR

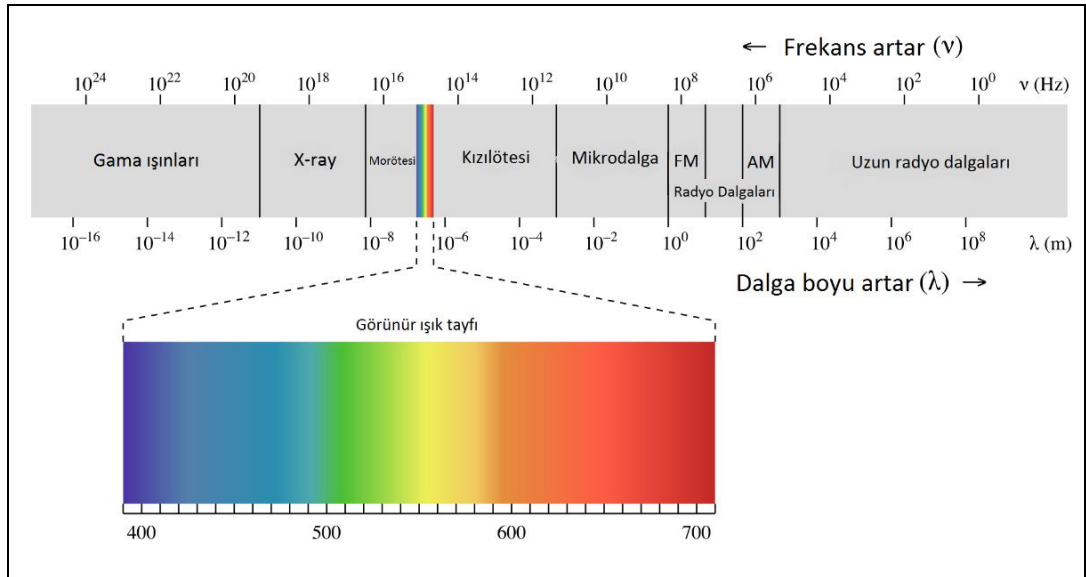
2.1. Elektromanyetik Enerji ve Elektromanyetik Spektrum

Uzaktan algılama disiplinde yeryüzündeki mevcut nesnelerin tanımlanması için nesnelerden yansıyan veya yayılan enerjinin miktarının belirlenmesi gerekmektedir. Bir enerji çeşidi olan ışık, güneşten gelen elektromanyetik dalgalar şeklinde yeryüzüne ulaşarak birçok algılayıcının (pasif algılayıcılar) yansıyan ışınları kullanmasını sağlar. Elektromanyetik enerji uzaktan algılama bileşenleri olarak bilinen hedef, enerji kaynağı ve algılayıcı arasında bağlantıyı oluşturmaktadır. Enerji iş yapabilme yeteneği olup iletim, taşınım ve ışıınım ile bir noktadan başka bir noktaya aktarılır. Uzaktan algılamada ise ışıınım ile enerji aktarılması önemli bir yer tutar. Elektromanyetik enerji dalga ve tanecik (foton) olmak üzere iki özellik göstermektedir. Elektromanyetik alanın uzayda yayılması ile oluşan elektromanyetik dalgalar birbirlerine ve ilerleme doğrultusuna dik olan elektrik alanı (E) ile manyetik (M) alandan oluşmaktadır [Campbell and Wynne, 2011]. Şekil 2.1’de elektromanyetik dalga yapısı verilmiştir [Web 1, 2021]. Bu dalgalar ışık hızında ($c = 3 \times 10^8 \text{ m/sn}$) hareket ederler ve dalga özelliği frekans (ν) ile dalga boyu (λ) kavramlarıyla tanımlanır. Dalga boyu iki dalga tepesi arasındaki mesafe olarak tanımlanırken, frekans ise belirli bir zamanda bir noktadan geçen dalga sayısı olarak tanımlanır.



Şekil 2.1: Elektromanyetik dalgalar.

Mutlak sıfırın (0 Kelvin, $-273,15\text{ }^{\circ}\text{C}$) üstünde sıcaklığa sahip olan bütün maddeler, farklı dalga boyuna sahip elektromanyetik dalgalar üretmektedirler. Üretilen dalga boylarının aralıklarının oluşturduğu yapı ise elektromanyetik spektrum olarak bilinir ve dalga uzunluğu nanometre derecesinden kilometreye kadar uzanmaktadır. Şekil 2.2’de elektromanyetik spektrumun dalga boyu aralıkları gösterilmiştir [Web 2, 2021]. Genel ifadeyle görünür bölge olarak adlandırılan aralık, insan gözünün algılayabildiği ve spektral duyarlılığı olan bölgedir. Genel aralık olarak görünür bölge $0,4\text{ }\mu\text{m}$ ile $0,7\text{ }\mu\text{m}$ dalga boyu aralığını temsil etmektedir. Görünür bölgede $0,4$ ile $0,5\text{ }\mu\text{m}$ dalga boyu arasına denk gelen bölge mavi, $0,5$ ile $0,6\text{ }\mu\text{m}$ dalga boyu arasına denk gelen bölgeye yeşil ve $0,6$ ile $0,7\text{ }\mu\text{m}$ dalga boyu arasına bölgeye kırmızı renge karşılık gelen dalga boylarıdır. Morötesi olarak adlandırılan bölge mavi bandın solunda yer alırken, görünür bölgenin bittiği yer olan kırmızı dalga boyuna bitişik bölge kızılötesi olarak adlandırılır. $0,7$ ile $1,3\text{ }\mu\text{m}$ dalga boyu arasında bulunan bölge yakın kızılötesi, $1,3$ ile $3\text{ }\mu\text{m}$ dalga boyu arasında bulunan yer orta kızılötesi, son olarak 3 ile $14\text{ }\mu\text{m}$ dalga boyu arasında kalan yer termal (uzak) kızılötesi olarak adlandırılmaktadır. Bu bölgelerin dışında kalan 1 mm ile 1 m aralığına sahip daha büyük dalga boylarının oluşturduğu bölge mikrodalga bölgesi olarak tanımlanmaktadır.



Şekil 2.2: Elektromanyetik spektrum.

Enerjinin ışıma ile taşınması olarak tanımlanan elektromanyetik radyasyonun, dalga teorisi ile açıklanan karakteristiğine rağmen enerji ve nesneler arasında meydana gelen ilişki kuantum teorisiyle tanımlanır. Bu teoriye göre enerji sürekli dalgalar halinde aynı

hız ve doğrultuda yayılma gösteren küçük tanecikler (fotonlar) olarak düşünülür. Bu durumda bir fotonun enerjisi 2.1’de verilen eşitlik yardımı ile hesaplanır ve eşitlikteki h Planck sabitini, c ışık hızını ve λ ise dalga boyunu temsil etmektedir.

$$E = h \frac{c}{\lambda} \quad (2.1)$$

Eşitlikten yola çıkılarak enerji ile frekans arasında doğru orantı, dalga boyu ile ters orantılı bir ilişki olduğu görülmektedir [Chuvieco, 2009]. Başka bir ifade ile dalga boyunun artmasına bağlı olarak taşınan enerji miktarının azalması söz konusudur. Uzaktan algılama uygulamalarında yeryüzü için yapılan çalışmalar ele alınırsa, mikrodalga gibi uzun dalga boyu özelliği gösteren bölgeden elde edilen enerji kızılötesi gibi daha kısa dalga boyundan elde edilecek enerjiden daha az olacaktır.

2.2. Uzaktan Algılamada Çözünürlük Kavramı

Uzaktan algılanan görüntülerin birbirinden farklılık göstermesini sağlayan bazı temel özelliklere mevcuttur. Bu farklılığı sağlayan dört tane çözünürlük kavramı vardır ve bunlar; mekânsal çözünürlük, spektral çözünürlük, radyometrik çözünürlük ve zamansal çözünürlüktür [Mather and Koch, 2011]. Mekânsal çözünürlük görüntüde bulunan nesnelerin ayırt edilebilme özelliğidir. Başka bir ifade ile uydu görüntüsünde detayların görünmesi algılayıcının mekânsal çözünürlüğü ile alakalıdır. Her bir pikselin temsil ettiği yeryüzü alanı algılayıcı için anlık görüş açısı (IFOV) ile bağlantılı olarak belirlenir. Uydu yüksekliğinin az olması daha küçük IFOV’a karşılık gelirken, yüksekliğin fazla olması daha büyük IFOV anlamına gelmektedir. Kısaca piksel boyutu ve mekânsal çözünürlük arasında doğrudan bir ilişki söz konusudur.

Spektral çözünürlük, herhangi bir algılayıcının elektromanyetik spektrum için kaydedebildiği dalga boyu aralığının yanında algılama yapılan bant sayısı ile tanımlanmaktadır. Bu çözünürlük çeşidi bant sayısı ve bant genişliği ile açıklanır. Bitki örtüsü ile su benzeri farklı arazi tipleri belirli dalga boyu aralıklarında ayırt edilebilirler. Bu sebeple yeryüzündeki nesnelerin veya arazi türlerinin belirlenmesi için spektral çözünürlükten yararlanılmaktadır. Sınıflandırma işleminde spektral

özünürlük diğ r özünürlük türlerine göre daha önemli bir yer tutmaktadır [Legleiter et al., 2002].

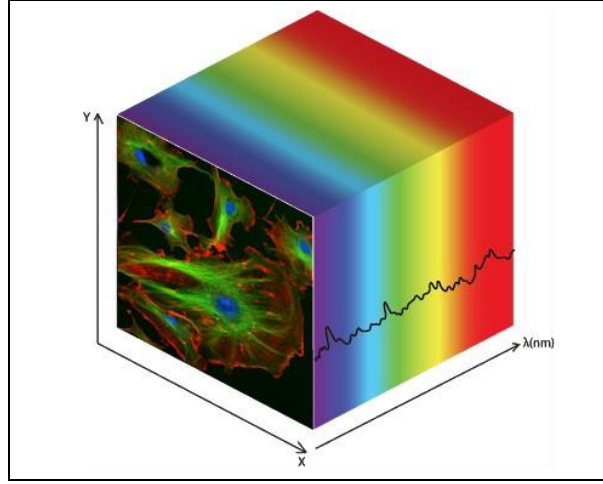
Bir diğ r özünürlük kavramı olan radyometrik özünürlük ise her bir pikselin sahip olduėu ışımaya şiddeti ile alakalıdır. Başka bir ifadeyle görüntü sağlayan algılayıcılar için gelen ışınımaya baėlı olarak enerjideki deėişikliklerin ayırt edilebilmesi olarak da tanımlanabilir. Görüntünün yapısı piksellerin yapısı ile alakalı iken, görüntü adına gerçek bilgi radyometrik özünürlük ile saėlanır. Nesnelerin tanımlanmasında ve ayırt edilmesinde kullanılan radyometrik özünürlük kavramı toplam renk tonu sayısına karşılık gelmektedir.

Son olarak zamansal özünürlük kavramı ise aynı bölge için hangi sıklıkta görüntü elde edildiėini ifade etmektedir. Uyduların yörüngedeki karakteristiėine baėlı olarak uydular arasında farklılık göstermektedir. Zamansal özünürlükte farklı zamanlarda ve zaman aralıklarında veri saėlanması mümkündür. Bu nedenle bulutsuz görüntü elde etmek, deėişim analizi alışmalarında kullanmak, afet olaylarını analiz etmek, tarım arazilerinde ürün gelişimlerinin zamana göre gelişmesini incelemek gibi birçok uygulamada zamansal özünürlükten yararlanılmaktadır.

2.3. Hiperspektral (Yüksek Boyutlu) Görüntüler

Uzaktan algılama alanındaki önemli gelişmelere baėlı olarak ortaya ıkan hiperspektral görüntü kavramı çok fazla banttan oluşan görüntüleri tanımlamaktadır. Bu verilerin elde edilmesini saėlayan hiperspektral alıcıların çok bantlı görüntü saėlama özelliėinin yanı sıra öne ıkan diğ r önemli özelliėi ise bantların ne kadar dar ve birbirleriyle bitişik olduklarıdır [Shippert, 2004]. Şekil 2.3'te hiperspektral bir görüntünün gösterim şekli verilmiştir [Web 3, 2021]. Şekilden görüldüėü üzere hiperspektral görüntüler ilk iki boyutu (X, Y) yersel bilgiyi gösteren, diğ r boyutu ise spektral bilgiyi temsil eden toplam üç boyuttan oluşan verilerdir.

Algılayıcıların saėladığı hiperspektral görüntülerin yanında, multispektral görüntülerin spektral bantlarından türetilmiş veya dış kaynaktan edinilmiş yardımcı verilerin görüntülere eklenmesiyle de yüksek boyutlu verilerin elde edilmesi mümkün hale gelmiştir.



Şekil 2.3: Hiperspektral görüntü.

Hiperspektral görüntüler madencilik, jeoloji, tıp, ekoloji gibi birçok alanda kullanılabilen verilerdir. Özellikle bitki ıslahı, hassas tarım gibi bitki fizyolojisini detaylı analiz edebilmek için yapılan çalışmalarda kullanılır [Thomas et al., 2018]. Bu tür veya farklı çalışmalarda tercih edilmesinde bazı avantajlara sahip olmaları etkilidir. Özellikle insan gözünün algıladığı görünür bölge olarak da adlandırılan dalga boyu aralığı dışında diğer dalga boyu aralığında da bilgi içerdiklerinden değerli veri türleridir. Bu nedenle hiperspektral görüntüler diğer uzaktan algılanan veri türlerine göre daha detaylı bilgi çıkarma potansiyeline sahiptirler. Ayrıca hiperspektral görüntü sağlayan algılayıcıların yanı sıra, multispektral olarak adlandırılan daha az bant sayısına sahip görüntülere çeşitli işlemlerin uygulanmasıyla veya bu görüntülere yardımcı verilerin eklenmesiyle de yüksek boyutlu görüntülerin elde edilebilmeleri mümkündür.

Bu görüntülerin sağladıkları avantajların dışında bazı dezavantajları da mevcuttur. Artan spektral bant sayısına bağlı olarak verilerin depolanması ve gerekli işlemlerin yapılması sorun teşkil etmektedir. Bu nedenle hiperspektral verileri analiz edebilmek için hızlı ve yüksek kapasiteli bilgisayarlar gerekir. Bu görüntülerin işlenmesi için zamana ihtiyaç duyulur ve işlenmeleri maliyetlidir. Hiperspektral görüntülerin bir diğer dezavantajı gereksiz ve tekrarlı bilgi içerebilmeleridir. Ayrıca spektral bantlar birbirine komşu ve dar aralıkta olduklarından aralarında meydana gelen korelasyon da hiperspektral görüntülerin analiz edilmesinde karşılaşılan başka bir sorundur. Hiperspektral görüntülerde boyutsallık sorunu olarak tanımlanan durum verilerin analiz edilmesini olumsuz olarak etkilemektedir [Hughes, 1968]. Bu

bağlamda arazi örtüsü/arazi kullanımı tematik harita üretme işlemi ele alındığında, hiperspektral görüntülerden yararlanabilmek için sınıflandırma işlemine hangi bantların dahil edileceği, hangilerinin sınıflandırma işleminde kullanılmaması gerektiği, kullanılacak bantların sayısı ve bant kombinasyonlarının belirlenmesi son derece önemli bir araştırma konusudur. Büyük boyutlu verilerin işlenmesi için üretilmiş farklı özellik seçim yöntemleri mevcuttur. Bu yöntemler hiperspektral görüntüler içinde kullanılabilir fakat birçok özellik seçim yönteminin varlığı hangi yöntemin kullanılacağı, hangisinin daha başarılı sonuç verdiği kullanılan metodun karmaşıklığına, veri setine ve problemin kendine has özelliklerine göre değişebilmektedir.

2.4 Görüntü Dönüşümleri

Görüntü dönüşümü, temel olarak multispektral veya gri tonlamalı görüntülerdeki bilgi içeriğinin bir kısmının ya da tamamının daha anlamlı bir şekilde yeniden düzenlenmesi işlemidir [Mather and Koch, 2011]. Görüntüde bilgi içeriğinin tamamını veya bir kısmı belirli işlemler ile yeniden ifade edilir. Uzaktan algılama uygulamalarında da yaygın şekilde kullanılan görüntü dönüşümleri belirli özelliklerin mevcut görüntüden daha fazla vurgulanması işlemidir. Başka bir ifadeyle görüntü dönüşümünün temel amacı görüntüde ilgilenilen özelliklerin belirgin hale getirilmesidir. Uzaktan algılanmış görüntüler için sınıflandırma işleminde kullanılacak yardımcı veriler, görüntü dönüşümlerinin ya da filtrelerin kullanılarak elde edilmesi mümkündür. Uzaktan algılamada kullanılan çeşitli görüntü dönüşüm yöntemleri vardır. Bunlara gri düzey eş oluşturma matrislerinden hesaplanan doku özellikleri, temel bileşenler analizi, ton-doygunluk-yoğunluk renk uzay dönüşümü, çeşitli işlemler ile spektral bantlardan hesaplanan spektral indisler ve morfolojik filtrelerin kullanılmasıyla elde edilen özellikler örnek olarak gösterilebilir.

2.4.1 Doku Özellikleri

Spektral ve mekânsal özellikler, görüntüler görsel olarak yorumlanırken dikkate alınan iki temel özelliktir. Gri tonlu görüntülerde, tonlamada meydana gelen değişiklikler spektral özellikleri tanımlanırken görüntü üzerindeki ton değişikliklerin

dağılımı mekânsal özellikleri tanımlamaktadır. Mekânsal özellikler görüntü üzerinde ilgili nesnede meydana gelen ton değişiklikleri ile görüntüdeki nesneler arasında meydana gelen ton değişikliklerini inceler [Tso and Mather, 2009]. Doku düzenli bir frekansta tekrar eden desen kombinasyonuna karşılık gelen ifadedir ve görüntü üzerinde ilgili nesnelerin tanımlanması için kullanılır. Doku özellikleri birbirine komşu olan iki piksel arasındaki uzaklığın ve açılarının esas alınmasıyla elde edilen gri düzey eş oluşum matrisleri ile hesaplanır [Haralick et al., 1973]. Belirli mesafe ve yöndeki piksel çiftleri arasında i elemanından j 'ye geçen olasılıkları ifade eden gri düzey eş oluşum matrisleri hesaplanırken genel olarak 3x3 veya 5x5 pencere boyutu kullanılır. Görüntü de 0° , 45° , 90° ve 135° yönlerinde desen özelliklerine göre analiz yapılır. Zıtlık, ortalama, varyans, homojenlik, farklılık, entropi, açısız ikinci moment ve korelasyon gri düzey eş oluşum matrislerinden hesaplanan doku özellikleridir. Zıtlık referans alınan piksel ile komşusu arasındaki yoğunluk veya gri düzey değişikliklerin ölçüsü olarak tanımlanmaktadır. Ortalama, pencere içerisinde bulunan piksellerin tamamını temsil eden ortalama gri ton değerini temsil eder. Diğer bir doku özelliği olan varyans ise referans piksel ve komşu pikseller arasındaki farkın dağılımını ifade eder. Homojenlik doku özelliği ise görüntüde farklı bölgeler için benzerliği açıklar dolayısıyla görüntüdeki ögeler aynı olduğunda en yüksek değere sahip olur. Başka bir doku özelliği olan farklılık görüntünün gri seviye farkı dağılımının ortalamasının bir ölçüsüdür. Entropi piksel çiftleri arasındaki düzensizliği temsil eden kavramdır. Diğer doku özelliği olan açısız ikinci moment görüntüdeki piksellerin tekdüzeliğini ölçmeye yarar. Son olarak korelasyon ise gri düzey değerlerde doğrusal bağımlılığın ölçüsüdür ve bir referans piksel ile komşusu arasında nasıl bir ilişki olduğu ile ilgilidir. Söz konusu doku özellikleri ve formülleri Tablo 2.1'de eşitlikler ile nasıl hesaplandıkları ifade edilmiştir. Tabloda i satır sayısını ve j ise sütun sayısını temsil ederken, $P_{i,j}$ olasılık yoğunluk fonksiyonunu temsil etmektedir. Ayrıca tabloda yer alan σ standart sapmayı, μ ortalamayı N ise görüntüde bulunan gri seviyelerin sayısını temsil etmektedir.

Tablo 2.1: Doku özellikleri ve formülleri.

Doku Özelliği	Formülü
Zıtlık	$\sum_{i,j=0}^{N-1} P_{i,j}(i+j)^2$
Ortalama	$\sum_{i,j=0}^{N-1} P_{i,j}/N^2$
Varyans	$\sigma^2_{i,j} - \sum_{i,j=0}^{N-1} P_{i,j}(i,j - \mu_{i,j})$
Homojenlik	$\sum_{i,j=0}^{N-1} P_{i,j}/(1 + (i+j)^2)$
Farklılık	$\sum_{i,j=0}^{N-1} P_{i,j} i-j $
Entropi	$\sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j})$
Açısal ikinci moment	$\sum_{i,j=0}^{N-1} P_{i,j}^2$
Korelasyon	$\sum_{i,j=0}^{N-1} P_{i,j} \left((i - \mu_i)(j - \mu_j) / \sqrt{(\sigma^2_i \sigma^2_j)} \right)$

2.4.2 Temel Bileşenler Analizi

Birbirleri ile korelasyonlu olan verilerin değerlendirilmesinde temel bileşenler analizi literatürde en çok tercih edilen yöntemlerden biridir [Jing and Panahi, 2006]. Temel bileşenler analizi kullanılarak orijinal veriden daha çok içeriğe sahip yeni veriler türetilir. Korelasyon iki değişken arasında gerçekleşen doğrusal ilişki olarak ifade edilir. Uzaktan algılanmış görüntülerde ise birbirine yakın olan spektral bantlar arasında korelasyon mevcuttur [Mather and Koch, 2011]. Bantlar arasındaki korelasyon gereksiz bilgi tekrarını ifade eder ve genellikle gerçek veri boyutu mevcut bantların sayısından azdır. Görüntülere temel bileşen analizi uygulanmasıyla mevcut görüntüden daha zengin ve depolamak için daha düşük boyutta görüntüler elde edilebilmektedir [Richards and Jia, 1999]. Analizin temel amacı bantlarda bulunan bilgileri tek bir görüntüde toplayarak, zengin içeriğe sahip ve korelasyonsuz yeni bir

görüntü oluşturmak ve görüntünün yorumlanmasını kolaylaştırmaktır. Bu işlem gerçekleştirilirken veri kaybı en düşük seviyede tutulmaktadır.

Temel bileşen analizinde yeni bir koordinat sistemi oluşturularak değerlendirilmeye alınan veri seti dikey eksenin oluşturduğu düzleme, matematiksel izdüşümü işleminin gerçekleştirilmesi söz konusudur. Üretilen yeni koordinat sistemi birbirlerine dik eksenlerden oluşmaktadır. Bu durum söz konusu eksenlerin arasında korelasyonu ortadan kaldırarak dönüştürmenin temel esasını oluşturmaktadır. Bu sayede uzaktan algılama uygulamalarında aralarında korelasyonlu ilişki bulunan bantların spektral anlamda karmaşıklığı giderilir. Temel bileşen analizi hesaplamak için iki adım takip edilir. İlk adım 2.2 eşitliği kullanılarak görüntüde bulunan tüm bantlar için varyans-kovaryans matrisinin bulunmasıdır.

$$Cov(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \quad (2.2)$$

Eşitlik 2.2’de x_i spektral banttaki gri değerini, $\bar{x}$ bandın gri değerlerinin ortalamasını, işleme alınan diğer bandın gri değeri y_i , söz konusu bandın ortalama gri değerleri $\bar{y}$ ile gösterilirken, n toplam pikseli ve $Cov(x, y)$ değerlendirilen bantlar arasındaki kovaryans değerini temsil etmektedir.

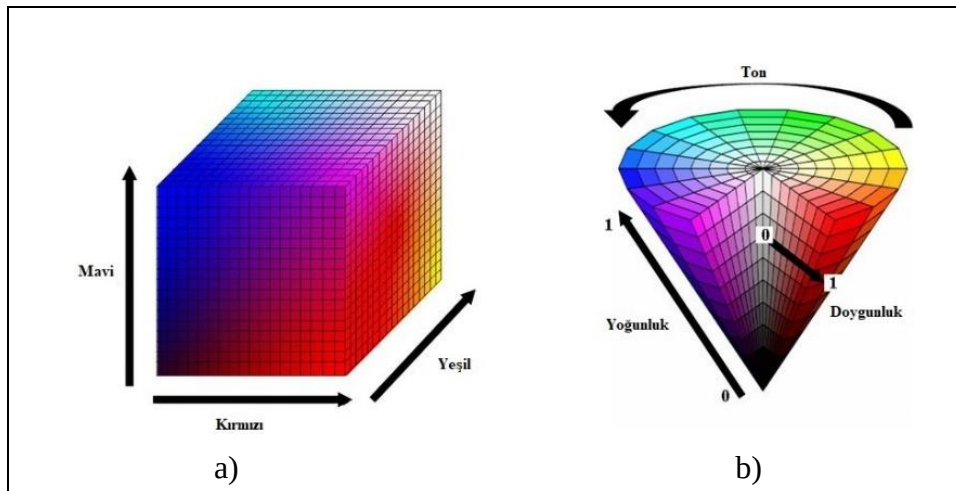
Diğer işlem adımı ise, elde edilen kovaryans matrisinin öz değerleri ile öz vektörlerinin bulunmasıdır. Vektörler yönü ve büyüklüğü olan niceliklerdir. Öz vektörler ise temel bileşenin yönelimi hakkında bilgi sağlamaktadırlar. Başka bir ifadeyle elde edilen özellik uzayının yönünü belirten kavramdır. Öz değerler ise temel bileşenlere ait eksen uzunluklarını ifade eder ve doğal olarak doğrusal bir şekilde dönüştürüldüklerinde öz vektörlerin artmasını ve azalmasını sağlayan sabitlerdir. Eşitlik 2.3’te A ($n \times n$) bir matris, v ($n \times 1$) sıfırdan farklı bir vektör ve λ sayısal bir büyüklüktür. Burada v öz vektör, λ ise öz değer olarak adlandırılmaktadır. Daha düzenli hale getirilmiş olan Eşitlik 2.4’te, öz vektörün (v) sıfır olmadığı durumda denklemin tek bir çözümü olduğu matematiksel ifade ile görülmektedir. Elde edilen eksen uzunlukları ilgili bileşenin içerdiği bilgiyi temsil eder. Temel bileşenler analizi hesaplama çıktıları, ilk bileşenin en yüksek bilgi seviyesini içerdiğini ve genellikle ilk üç bileşenin mevcut bilginin büyük bir kısmını içerdiğini göstermektedir.

$$A \times v = \lambda \times v \quad (2.3)$$

$$|A - \lambda \times I| = 0 \quad (2.4)$$

2.4.3 Ton-Doygunluk-Yoğunluk Renk Uzay Dönüşümü

Renk çeşitliliğinin fazlalığı nedeniyle renkleri temsil edebilmek için renk uzayı kavramı ortaya çıkmıştır. Her renk uzayının kendine has bir yapısı vardır ve renk uzayları matematiksel modellerdir. Renklerin elde edilebilmeleri için üç değişkene ihtiyaç vardır bu yüzden renk uzayları üç boyutlu olarak tanımlanırlar. En çok bilinen renk uzayı kırmızı, yeşil ve mavi (RGB) ana renklerden oluşur. Şekil 2.4'te RGB renk uzayı birim küp olarak temsil edildiği görülmektedir. Bu renk uzayında temsil edilen üç ana rengin çeşitli kombinasyonları insan gözü tarafından algılanabilmektedir. RGB renk uzayında, 8 bite sahip olan bir görüntü için değerler 0 ile 255 aralığındadır [Mather and Koch, 2011]. Uzaktan algılama uygulamalarında yardımcı veri oluşturulması amacıyla RGB renk uzayından farklı renk uzaylarına dönüşümler mümkündür. Bu uzaylardan biri olan HSI renk uzayı, altıgen bir koni ile temsil edilir ve renkleri yoğunluk-ton-doygunluk olarak üç bileşenle tanımlanır [Chernov et al., 2015].



Şekil 2.4: a) RGB renk uzayı, b) HSI renk uzayı.

Yoğunluk, rengin parlaklığını, yani rengin beyaz ile ne kadar seyreltildiğini göstermektedir. Ton, rengi oluşturan ışığın baskın veya ortalama dalga boyunu,

doygunluk ise rengin canlılığını yani saflık derecesini tanımlamaktadır. RGB renk uzayından HSI renk uzayına dönüşümün sonucu olarak RGB renk uzayının temsil ettiği 0 ile 255 arasındaki değerler HSI renk uzayı modelinde yoğunluk ve doygunluk olarak 0 ile 1 arasında değerler alır. Ayrıca ton 0° ile 360° arasında açısal değerlere sahiptir. Uzaktan algılama yöntemi ile elde edilen görüntüler kırmızı, yeşil ve mavi bantların çeşitli matematiksel işlemlerle beraber kullanılarak ton-doygunluk-yoğunluk değerleri hesaplanmaktadır. Eşitlik 2.5, Eşitlik 2.6 ve 2.7’de verilen R kırmızı bandı, G yeşil bandı, B mavi bandı, I yoğunluk değerini, S doygunluk değerini ve son olarak H ise ton değerini temsil etmektedir.

$$I = \frac{R + G + B}{3} \quad (2.5)$$

$$S = 1 - \frac{3[\min(R, G, B)]}{(R + G + B)} \quad (2.6)$$

$$H = \cos^{-1} \frac{1/2 \times ((R - G) + (R - B))}{((R - G)^2 - (R - B)(G - B))^{1/2}} \quad (2.7)$$

2.4.4 Spektral İndisler

Spektral bantların matematiksel eşitliklere bağlı olarak birbirleri ile oranlanması sonucu elde edilen spektral indisler uzaktan algılama uygulamalarında oldukça yaygın bir kullanıma sahiptir. Ayrıca, spektral indisler farklı bantların yansıma değerlerine bağlı olarak doğrusal kombinasyonlar ya da oranlarını içeren matematiksel dönüşümler olarak da tanımlanabilirler. Spektral indisler, görüntüde farklılık gösteren topoğrafyadan kaynaklanan parlaklık değişimini ortadan kaldırarak verinin renk içeriğinin vurgulanmasını sağlarlar [Lillesand et al., 2015]. Görüntülerde vurgulanmak istenen topoğrafik özelliğe göre farklı indisler kullanılabilir. Uzaktan algılamada, spektral indisler yardımcı veri olarak kullanıldığında sınıflandırma doğruluğuna olumlu etki ettiği gözlemlenmiştir. Nesnelerin spektral imzaları kendilerine özgü bir özelliktir. Bu nedenle uzaktan algılanmış görüntülerde parlaklık değişimlerinin

engellediği ince spektral ayrımların belirlenmesinde katkı sağlayan indisler vurgulanmak istenen nesneye göre bant oranları olarak kullanılmaktadır. Bitkinin, suyun ve toprağın yansıtım özellikleri incelendiğinde, yakın kızıl ötesi dalga boyu bölgesinde birbirlerinden önemli ölçüde farklılık gösterdikleri bilinmektedir. Bu sebeple çalışma konusuna bağlı olarak spektral bant üretmek için hangi dalga boyu aralığının seçileceği önem arz etmektedir. Örneğin, sağlıklı bitkiler açısından düşünülürse yakın kızıl ötesi (NIR) bandın kırmızı banda oranı yüksektir. Yakın kızıl ötesi bandın yansıtma değerinin azaltılması ve kırmızı bandın yansıtma değerinin artırılması durumunda görüntüde baskılanmış bitki örtü türlerinin ortaya çıkarılması sağlanabilir. Böylece yakın kızıl ötesi bandın kırmızı banda oranlanması görüntüde olan bitki örtüleri arasındaki ayrımın yapılabilmesine olanak sağlamaktadır. Böyle bir oranlamadan biokütle veya bitki örtüsü sağlığının belirlenmesi gibi birçok çalışmada yararlanılmaktadır.

Literatürde farklı çalışmalara uygun olarak kullanılan birçok spektral indis mevcuttur. Uzaktan algılanmış görüntülerde spektral indisler vejetasyon, su, yanmış alan, yerleşim yeri ve toprak gibi farklı uygulamalarda başarılı bir şekilde kullanılmaktadır. Vejetasyon ile ilgili çalışmalar diğerlerine göre daha çok yapılmaktadır. Bu çalışmalar yıllık ürün tahmini, akıllı tarım ve ormancılık gibi alanlarda sıklık göstermektedir. En bilinen vejetasyon indislerinden olan Normalleştirilmiş Fark Oran İndisi (NDVI), Eşitlik 2.8’de ki gibi yakın kızılötesi ve kırmızı bantları kullanılarak hesaplanır. En temel indis olan NDVI hesaplama sonucunda -1 ile 1 aralığında değişen değerlere sahip olduğu bilinmektedir. NDVI yeşil bitki örtüsü için 1 değerine yaklaşan değerler alırken su ve kar gibi arazi yüzeylerinde eksi, zayıf bitki örtüsü ve çıplak toprak gibi arazi yüzeylerinde sıfıra yaklaşan değerler almaktadır [Hatfield et al., 1985]. Bitki analizi veya tarım uygulamalarında sadece NDVI kullanılmasının yanı sıra farklı spektral indislerin eklenmeleriyle veri setinin genişletilmesi ve daha gerçekçi bir analiz yapılabilir.

$$NDVI = \frac{NIR - Kırmızı}{NIR + Kırmızı} \quad (2.8)$$

2.4.5 Morfolojik Filtreler

Morfolojik işlemler, görüntülerde ilgilenilen nesne veya bölge özelliklerinin belirgin hale getirilmesi veya görüntüde mevcut olan gürültünün ortadan kaldırılması amacıyla matematiksel işlemler ile elde edilen filtrelerdir. Bu işlem çoğunlukla 0 ve 1 piksel değerlerinden oluşan görüntüler (binary) için kullanılmasının yanında gri düzey görüntüler içinde kullanılmaktadır [Moore et al., 2006]. Morfolojik filtre işleminde yapısal eleman (Structuring Element) olarak adlandırılan matrisler kullanılmaktadır. Bu matrisler görüntü üzerinde dolaştırılarak elde edilmek istenen morfolojik işleme göre görüntüdeki pikseller ile karşılaştırılır ve sonuç görüntüsü elde edilir. Yapısal elemanlar farklı şekillerde veya büyüklüklerde olabilir, bir merkez noktaları mevcuttur ve görüntüde bulunan her piksel bu nokta ele alınarak işleme tabi tutulur [Acar and Bayram, 2009]. Morfolojik operatörler genişleme, aşınma, açılma ve kapanma işlemleri olmak üzere dört farklı şekilde gerçekleştirilebilir.

Morfolojik genişleme, görüntü üzerinde dolaştırılan yapısal elemanına göre görüntüde bulunan piksellerde genişletme ve büyütme işleminin gerçekleştirilmesidir. Koyu tona sahip olan pikseller daha parlak pikseller haline getirilir ve sonuç olarak nesnelerin boyutları büyümüş olur. Eşitlik 2.9'da A görüntüsüne B yapısal eleman uygulanarak elde edilen genişleme işlemi $(A \oplus B)$ şeklinde gösterilmiştir.

$$A \oplus B = \bigcup_{b \in B} A + b \quad (2.9)$$

Morfolojik operatörlerinden olan aşınma işlemi genişleme işleminin tersidir. Görüntüye uygulanan yapısal elemanına göre piksellerde küçülme meydana gelir. İşlem uygulandıktan sonra görüntüdeki nesneler küçülmüş olur. Eşitlik 2.10'de A görüntüsüne B yapısal eleman uygulanarak elde edilen aşındırma işlemi $(A \ominus B)$ ile gösterilmiştir.

$$A \ominus B = \bigcap_{b \in B} A - b \quad (2.10)$$

Açılma morfoloji operatörü, aşınma ve genişleme işlemlerinin birbiri ardına uygulanması ile elde edilir. Bu işlem sonucunda görüntüdeki nesneler arasında

bulunan boşluklar yapısal elemana bağlı olarak temizlenir. Eşitlik 2.11’de A görüntüsü B yapısal elemanı kullanılarak elde edilen genişleme ve aşındırma işlemleriyle elde edilen açma işlemi $(A \circ B)$ ile gösterilmiştir.

$$A \circ B = (A \ominus B) \oplus B \quad (2.11)$$

Kapanma morfoloji operatörü ise açma işleminin tam tersidir yani önce genişleme işlemi sonra aşınma işlemi kullanılarak elde edilir. Kapanma işlemi sonucunda görüntüdeki nesneler arasında bulunan boşluklar doldurulur. Eşitlik 2.12’de A görüntüsü B yapısal elemanı kullanılarak elde edilen genişleme ve aşındırma işlemleriyle elde edilen açma işlemi $(A \bullet B)$ ile gösterilmiştir.

$$A \bullet B = (A \oplus B) \ominus B \quad (2.12)$$

3. UZAKTAN ALGILAMADA SINIFLANDIRMA

Uzaktan algılanmış görüntülerin geniş kullanım yelpazesine bağlı olarak görüntülerin çalışma amacına uygun şekilde işlenmesi zaruri bir ihtiyaç olarak karşımıza çıkmaktadır. Görüntülerde bulunan doğal veya yapay nesneler en küçük görüntü elemanı olan pikseller ile temsil edilmektedir. Söz konusu piksellerin temsil ettiği arazi sınıfına atanması sınıflandırma işlemi denir [Lillesand et al., 2015]. Görüntü sınıflandırma işlemi piksellerin esas alındığı piksel tabanlı ve nesnelerin çalışmaya konu olduğu nesne tabanlı sınıflandırma olmak üzere iki şekildedir. Farklı yöntemlerinin varlığıyla beraber sınıflandırma işlemi uydu görüntülerin yorumlanmasında önemli bir yer tutar [Radhika and Varadarajan, 2016].

Uydu görüntülerinde sınıflandırma işlemi, piksellerin veya görüntü nesnelerini veri değerlerine göre gruplandırılma işlemi olarak tanımlanabilir [Sisodia et al., 2014]. Başka bir ifadeyle, piksel tabanlı sınıflandırma işlemi, spektral özelliklerin temsil edildiği piksellerin belirlenmiş sınırlı sayıda olan sınıfa atanması olarak da açıklanabilir. Uzaktan algılanmış bir görüntünün sınıflandırma işlemindeki temel esas farklı spektral yansımalara sahip piksellerin bu özelliklerinden yararlanılmasıdır.

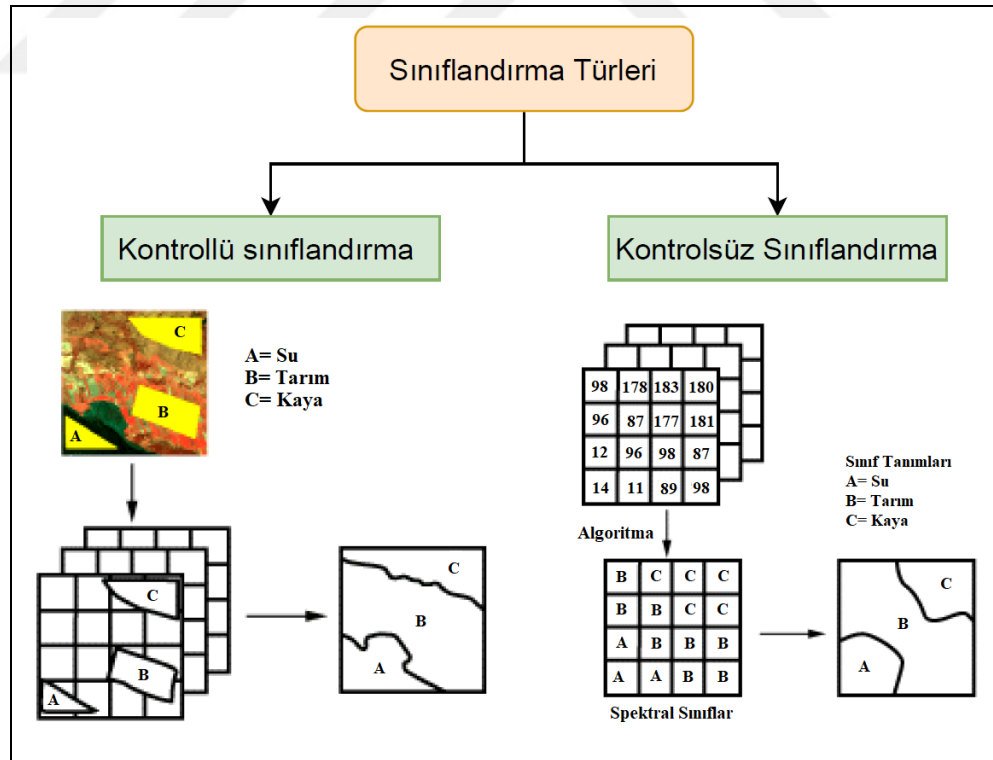
Görüntülerin sınıflandırılmasıyla yeryüzüne ait farklı özelliklerin temsil arazi örtüsü/arazi kullanımını gösteren tematik haritalar üretilmiş olur. Tematik haritalarda pikseller belirli sınıflar ile temsil edildiğinden görüntülerin analiz edilmesi ve yorumlanması açısından kolaylık sağlamaktadır. Arazi örtüsü bölgede orman, tarım ve su gibi alanların dağılımı hakkında bilgi sağlarken arazi kullanımı arazi kullanımı ise bölgede insan faaliyetleri ile arazinin nasıl kullanıldığı hakkında fikir verir. Uzaktan algılamada sınıflandırma işlemi genellikle kontrollü ve kontrolsüz sınıflandırma olmak üzere iki şekilde yapılmaktadır.

3.1. Kontrollü ve Kontrolsüz Sınıflandırma

Kontrollü sınıflandırma, kimliği belirlenmiş piksellerin eğitim verisi olarak kullanılıp, belirsiz olanları bir sınıfa atama işlemidir [Campbell and Wynne, 2011]. Kontrollü sınıflandırma işlemini gerçekleştirebilmek için belirlenmiş sınıflara ait eğitim verisi kullanılır. Tercih edilen sınıflandırma yöntemine göre eğitim verisi

kullanılarak bir model oluşturulması söz konusudur. Modelin tüm görüntüye uygulanması ile piksellerin ait oldukları sınıflar belirlenmiş olur. Üretilen tematik haritalar değerlendirildiğinde kontrollü sınıflandırma, doğruluk ve sınıflandırma sonucunun güvenilirliğine açısından uzaktan algılama çalışmalarında en çok tercih edilen yöntemdir.

Görüntüde bulunan ve etiketi belli olmayan piksellerin eğitim verisi kullanılmadan bilinmeyen piksellerin belirlendiği çözüm yollarına kontrolsüz sınıflandırma denir [Lillesand et al., 2015]. Kontrolsüz sınıflandırmada, piksellerin parlaklık ve spektral özelliklerine göre gruplandırma işlemi gerçekleştirildiği için oluşturulan sınıflar (gruplar) spektral sınıflardır. Diğer bir ifadeyle, asıl amaç belirli spektral benzerliklere dayanan spektral grupları üretmektir denilebilir. Kısaca kontrolsüz sınıflandırma iki aşamadan oluşmaktadır. İlk aşama benzer spektral yansımaya sahip yeryüzü nesnelere göre veri kümelerine ayrılır. İkinci adımda ise elde edilen spektral sınıfların hangi arazi örtüsüne ait olduğu ise kullanıcı tarafından belirlenir. Şekil 3.1’de kontrollü ve kontrolsüz sınıflandırma yaklaşımlarına ait örnekler gösterilmiştir.



Şekil 3.1: Kontrollü ve kontrolsüz sınıflandırma.

İki sınıflandırma yöntemi karşılaştırıldığında bazı temel farklılıkların olduğu görülmektedir. Kontrolsüz sınıflandırma işleminde ayrıntılı bir ön bilgiye ihtiyaç duyulmamasının yanında sınıflandırma işlemi ile üretilen sonuçların anlamını yorumlamak için çalışma alanının bilgisine ihtiyaç duyulmaktadır. Ayrıca kontrolsüz sınıflandırma işlemlerinde uygulama aşaması için daha az zamana ihtiyaç duyulur. Kolay bir yöntem olmasıyla birlikte, özgün spektral sınıflar oluşturulmaktadır. Ancak elde edilen spektral sınıflar her sınıfı tanımlayamayabilir, sınıfların yorumlanması için zamana ihtiyaç duyulabilir ve düşük doğrulukta sonuç üretirler. Diğer yandan, kontrollü sınıflandırma işlemi gerçekleştirmek için çalışma alanı hakkında ayrıntılı bilgi, haritalanacak her sınıfın temsili örneklerini seçmek gerekir. Kontrolsüz sınıflandırma işlemi ile tanımlanan sınıflar kontrollü sınıflandırma ile üretilenlerden spektral açıdan daha homojen bir yapıda olmaktadır. Kontrolsüz sınıflandırma işleminde sınıf kimlikleri hakkında analist sınırlı bir kontrol düzeyine sahiptir. Kontrollü sınıflandırma işleminde belirli bir amaca ve arazi yüzeyine göre uyarlanmış sınıfların varlığı söz konusudur. Ayrıca analist kontrolsüz sınıflandırma gibi elde edilen sınıfları arazi yüzeyi ile eşleştirmek zorunda değildir çünkü bu işlem eğitim verilerinin seçilmesi sürecinde ele alınmıştır. Bu nedenle analizi yapan kişi sınıflandırmayı kontrol edebilir ve söz konusu eğitim verilerinin tekrardan kullanılabilmesi mümkündür. Kullanılacak olan eğitim ve sınıflandırma doğruluğunun belirleneceği test verilerinin toplanması zaman almaktadır ve arazi sınıfları ile belirlenmiş olan spektral sınıfların birbiriyle eşleşmeme durumuyla karşılaşılabilir. Bu bilgilerin ışığında yapılacak uygulamaya uygun olacak şekilde sınıflandırma yöntemi seçilir.

3.2. Sınıflandırma Algoritmaları

Uzaktan algılanmış görüntüleri sınıflandırmak için bugüne kadar çeşitli yöntemler geliştirilmiştir. İstatiksel tabanlı popüler sınıflandırma yöntemi olan en yüksek olasılık algoritması eğitim aşamasında, sınıflandırıcıya tanıtılan her sınıf için verilerin normal dağılımında olduğunu varsaymaktadır [Kavzoglu, 2017]. Söz konusu algoritma gibi geleneksel sınıflandırma tekniklerinin yanında, çeşitli ileri sınıflandırma algoritmaları da başarıyla uygulanmaktadır. Özellikle 1990'lı yıllarda yapay sinir ağı (YSA) yaklaşımları ile uzaktan algılanmış verilerin sınıflandırılması

alanında başarılı arařtırmalar yapılmıřtır [Benediktsson et al., 1993], [Yang, 1999]. İnsan beyninin alıřma ilkesi esas alınarak retilen YSA'lar, belirli bir ğrenme grevini yerine getirmek iin paralel hesaplama ve yksek hesaplama gc gerektirmektedir [Kavzoglu, 2001]. Geliřtirilen diğerk bir sınıflandırma algoritması olan destek vektr makineleri (DVM), iki sınıfı en iyi řekilde ayıran hiper dzlemi tanımlama yntemine dayanmaktadır [Vapnik, 1995]. Hiper dzlem yapısal riski en aza indirerek zlen bir optimizasyon problemi kullanarak elde edilir. İstatistiksel yaklařımların aksine, DVM algoritmasıyla verilerin olasılık dağılımı hakkında herhangi bir varsayım yapılmadan verilerdeki sınıflandırma hatası en aza indirilir. 2000'li yıllar ve sonrasında ise ok fazla ileri sınıflandırma yntemi geliřtirilmiř ve makine ğrenmesi alanında bir ığır aılmıřtır. Sz konusu geliřtirilen yntemlerden biri olan rastgele orman algoritması (RO), karar ağlarını temel alarak toplu bir ğrenme modeli oluřturma ilkesine dayanmaktadır [Breiman, 2001]. RO sınıflandırma yntemi birok alıřmada başarılı bir řekilde kullanılmaktadır. Teknolojinin ilerlemesiyle ivme kazanan alıřmalar neticesinde son yıllarda yeni sınıflandırma algoritmaları retilmiřtir. Ařırı gradyan artırma makineleri (XGB) sahip olduėu yksek performans ile başarılı bir řekilde uygulanan, karar ağı temelli yeni nesil bir sınıflandırma yntemidir [Chen and Guestrin, 2016]. Her yinelemeden sonra yeni bir sınıflandırma yeliğini tahmin edilir. XGB algoritması ile ařırı ğrenme (overfitting) probleminin giderilerek model doėruluėunun artırılmaktadır. Yakın zamanda geliřtirilen bir diğerk karar ağ tabanlı sınıflandırma metodu olan hafif gradyan artırma makineleri (LightGBM), yksek boyutlu veriler ile alıřılırken gerekli olan zaman maliyetini dřrmek iin nerilmiřtir [Ke et al., 2017]. LightGBM algoritmasında karar ağlarının dallanma řeklinin farklı olmasıyla diğerk gradyan artırma yntemlerinden ayrılmaktadır.

Geliřtirilen yeni metotlarla daha yksek sınıflandırma doėruluėuna sahip tematik haritaların elde etmek iin eřitli sınıflandırma yntemlerinin retildiėi grlmektedir. Uzaktan algılanmıř grntlerin sınıflandırılmaları iin kullanılan algoritmaların dayandıkları matematiksel esaslar ve alıřma mantıkları birbirlerinden farklılık gstermektedir. Bu nedenle her bir yntemin uygulanmasıyla hesaplanan genel doėruluk ve sergilemiř oldukları performanslar da farklılık gsterecektir.

3.2.1. Maksimum Olabilirlik Yöntemi

En yüksek olasılık sınıflandırıcısı, parametrik ve istatistiksel bir yöntem olarak uzaktan algılama uygulamalarında kullanılan en yaygın yöntemdir [Mustapha et al., 2010]. Bayes olasılık temeline dayanan bu yöntemle, önceden tanımlanmış olan arazi sınıflarına ait olan piksellerin olasılığı hesaplanır ve daha sonra olasılığın en yüksek olduğu arazi sınıfına atanır [Tso and Mather, 2009]. Genellikle uygulamalarda normal dağılım modeli kullanılmakta olup, her bir sınıfa ait eğitim verisinin normal dağılım sergilediği kabulü söz konusudur. Bu kabul ile birlikte özellikle optik uzaktan algılanmış verilerin modellenmesinde, daha sağlam algoritmaların meydana gelmektedir. Yöntemde normal dağılım tek değişkene sahip olan veriler için ortalama ile varyans parametreleriyle, çok değişkene sahip olanlarda ise ortalama vektörü ile kovaryans matrisi yardımıyla belirlenmektedir. Özellik uzayının (p) 1'den büyük olması halinde ortalama vektör ve kovaryans matris kullanılarak her bir kategorinin dağılımı hesaplanır. Dağılımın oluşmasını sağlayan matematiksel formül ikinci dereceden olduğu için özellik uzayında parabol, daire veya elips gibi eğriler oluşabilmektedir. En yüksek olabilirlik yönteminde Eşitlik 3.1'de Bayes teorisine göre, herhangi bir x örneğinin S_i sınıfına ait bir örnek olma olasılığı $P(S_i | x)$ ile ifade edilmiştir.

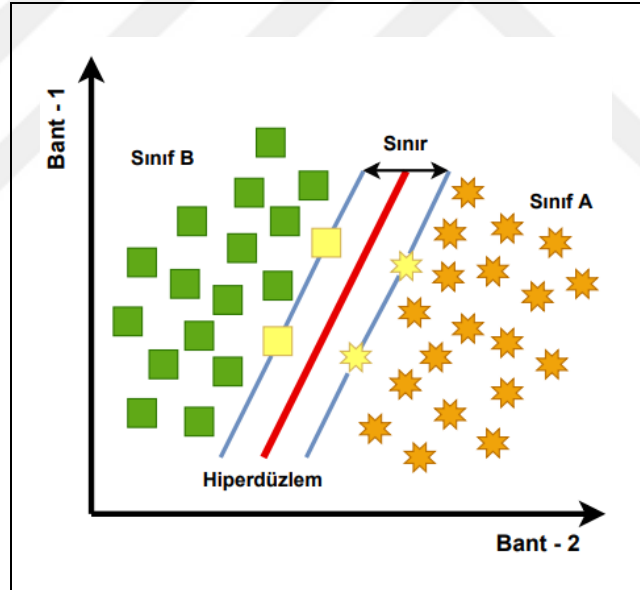
$$P(S_i | x) = \frac{P(S_i | x) P(S_i)}{P(x)} \quad (3.1)$$

3.2.2. Destek Vektör Makineleri

Destek vektör makineleri (DVM) birçok çalışmada olduğu gibi uzaktan algılama çalışmalarında da etkili şekilde kullanılmış ve üstün performans gösterdiği kanıtlanmıştır [Kavzoglu and Colkesen, 2009], [Pal and Mather, 2005], [Varma et al., 2016]. DVM istatistiksel öğrenme teorisine dayanan ve sınıfların en uygun şekilde ayrılmasını sağlayan bir sınıflandırma yöntemidir [Vapnik, 1995]. Ayrıca DVM parametrik olmayan bir sınıflandırıcı yöntemidir ve bu yöntem ile karar sınırlarının yerinin en iyi şekilde belirlenmesi amaçlanır. Parametrik sınıflandırmada amaç, her sınıfın tipik özellik uzay değerlerini veya dağılımlarını karakterize etmektir. DVM'de

ise sadece özellik alanında sınıflar arasındaki en uygun sınıra yakın olan eğitim örneklerine odaklanılır. Bu yöntemde en önemli ilke iki sınıfa ait olan piksellerin birbirlerinden en ideal şekilde ayırabilecek bir hiperdüzlemin (hyperplane) belirlenmesidir [Kavzoglu and Colkesen, 2009]. Şekil 3.2’de DVM’nin iki sınıf için doğrusal yapıda olan bir veri setinin sınıflandırılması işlemi gösterilmiştir. Şekilde A sınıfına ait olan pikseller turuncu yıldızlar şeklinde ve B sınıfına ait pikseller yeşil kareler şeklinde gösterilmiştir. Sınır olarak tanımlanan uzaklık mavi çizgiler arasındaki mesafeyi gösterir ve genişliğine göre sınıfların iyi ayrıldığı belirtmektedir. A ve B sınıfına ait piksellerin arasında maksimum uzunluğu gösteren yalnız bir tane sınır çizgisi vardır.

Her iki sınıfa ait mavi çizgilerin üstünde bulunan sarı pikseller destek vektörleri olarak tanımlanır ve yönteme ismini verir [Maxwell et al., 2017]. Kırmızı çizgi sınırın tam ortasından geçecek şekilde yerleştirilir ve bu çizgiye hiper düzlem adı verilir.



Şekil 3.2: Destek vektör makinelerinin iki sınıfı ayırmak için kullanımı.

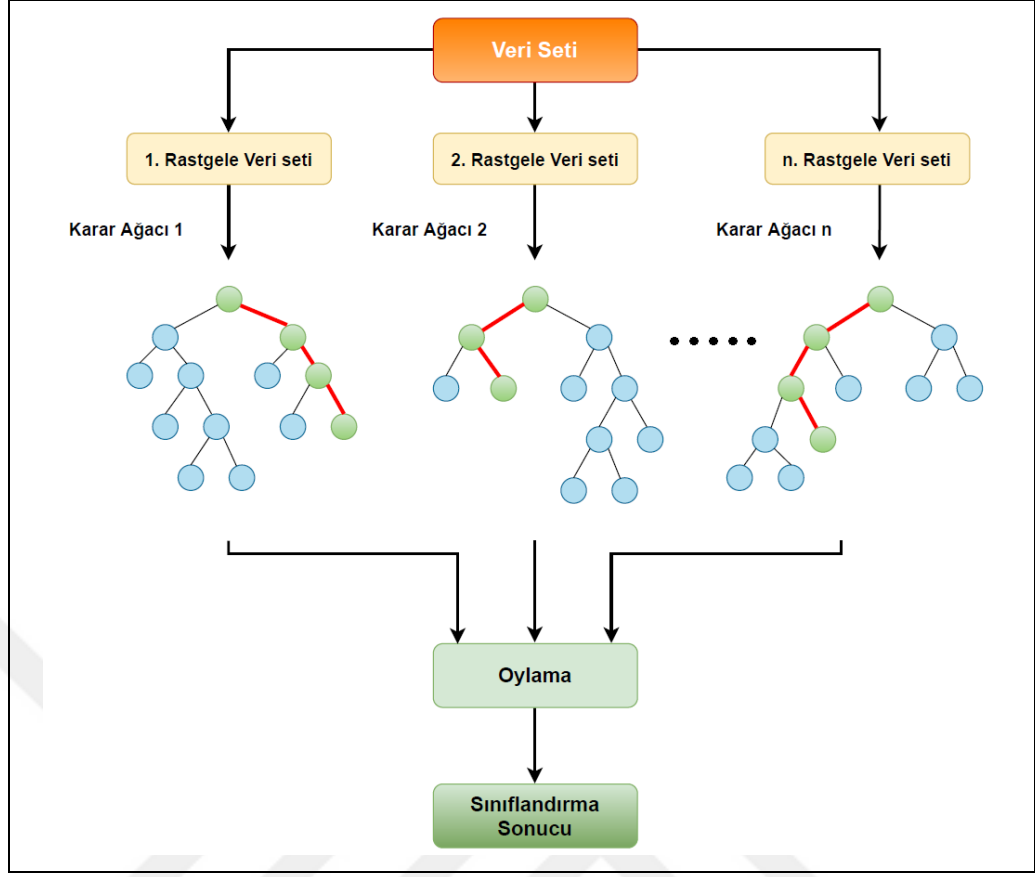
Sınıf etiketi belirsiz olan aday pikseller belirlenen hiper düzleme göre ilgili sınıfa atanır bu nedenle DVM sınıflandırma sürecinde hiper düzlem oluşturma önemli bir yeri vardır. Sınıfı belirlenecek aday piksel, şekilde gösterildiği gibi hiper düzlemin sağ tarafında yer alırsa A sınıfına, sol tarafında yer alırsa B sınıfına atanır. DVM’nin çalışma prensibinden anlaşılacağı üzere işlem sürecinde yalnızca destek vektörleri olarak tanımlanan pikseller dikkate alınır. Diğer piksellerin ise sınıflandırma işlemine

etkileri yoktur. Bu yüzden DVM ile daha az sayıda eğitim verisi kullanılmasına rağmen diğer sınıflandırma algoritmalarına göre üstün performans gösterebilmektedir [Kavzoglu and Colkesen, 2012]. DVM için farklı çekirdek tipleri kullanılabilir.

3.2.3. Rastgele Orman Algoritması

Topluluk sınıflandırma algoritmaları, tek bir sınıflandırıcı yerine bir dizi sınıflandırıcı oluştururlar ve tahminlerini oylama yöntemi ile gerçekleştiren sınıflandırma algoritmalarıdır. Toplu öğrenme algoritmalarından biri olan Rastgele orman (RO) algoritması sağladığı yüksek doğruluk ile uzaktan algılama alanında popüler bir hale gelmiştir [Belgiu and Drăgut, 2016], [Kavzoglu 2017]. RF, karar ağaçlarını kullanarak karar ağacı ormanı oluşturan bir sınıflandırma yöntemidir [Breiman, 2001]. Ormanda bulunan karar ağaçlarının eğitilmesi için orijinal eğitim veri setinden rastgele alt kümeler elde edilir. Oluşturulan bu alt kümelerin 2/3'ünden karar ağaç yapısını meydana getirmek için yararlanılır [Kavzoglu et al., 2020]. Eğitim verilerinden kalan 1/3'lük kısmı ise (Out of Bag, OOB) ağaç yapısının test edilmesi için kullanılmaktadır. Her bir karar ağacı için sınıflandırma sonucunda bir oy alması söz konusudur ve en çok oya sahip olanın tespit edilmesiyle ağaç yapısı belirlenmektedir.

Şekil 3.3 ile de temsil edilen RO algoritmasının dallanma ölçütlerinin seçiminde Gini indeksi metodu kullanılmaktadır. RO algoritması için önemli olan ve kullanıcı tarafından belirlenen iki parametre vardır. İlk parametre her düğümü bölmek için kullanılan örnek sayısı ve ağaç sayısıdır. RO algoritması kullanıcı tarafından belirlenen ağaç n tane ağaçtan oluşur. Yeni veri setleri için sınıflandırma işlemi sırasında, n sayıda belirlenmiş ağaçların hepsinden geçirilir. Bu ağaçlardan elde edilen n sayıda oydan en çok oya sahip olan sınıfın seçilmesiyle veri setine ait sınıf belirlenmiş olur



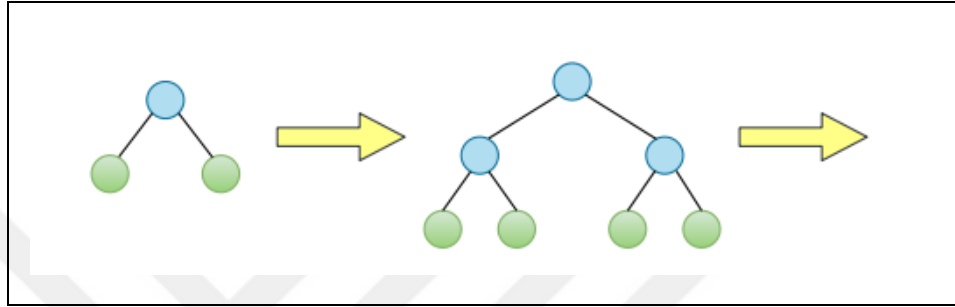
Şekil 3.3: Rastgele orman algoritması çalışma ilkesi.

3.2.4. XGBoost Algoritması

Aşırı gradyan artırma (XGB) son yıllarda sağladığı yüksek performans ile başarılı bir şekilde kullanılan, karar ağacı temelli yeni nesil bir makine öğrenme algoritmasıdır [Chen and Guestrin, 2016]. XGB makine öğrenmesi algoritmasının temel işlevi, her yinelemeden sonra yeni bir sınıflandırma üyeliğini öngörmesidir [Georganos et al., 2018]. Söz konusu algoritmanın amacı, eğitim aşamasında aşırı öğrenme (overfitting) probleminin giderilerek model doğruluğunun artırılmasıdır. Yöntemde kullanılan amaç fonksiyonu yüksek başarının elde edilmesini sağlar ve bu amaç fonksiyonu kayıp fonksiyonu ile düzenleştirme (regularization) kavramlarından oluşmaktadır [Üstüner ve diğ., 2020]. Kayıp fonksiyonu model ile yapılan her bir tahmin değeri ile gerçek değeri arasındaki farkı hesaplamaya yararken, aşırı öğrenme sorununun üstesinden gelinmesi düzenleştirme fonksiyonu ile sağlanır [Chen & Guestrin 2016]. Ayrıca düzenlilik terimi, model karmaşıklığını cezalandırarak daha sağlam bir model oluşmasını sağlamaktadır. Kullanılan bu düzenlilik fonksiyonu ile

gradyan artırma makinelerinden (GBM) ayrılmaktadır. Bu nedenle GBM yöntemine göre hızlıdır.

XGB model oluştururken optimize edilebilecek birçok parametreyi bünyesinde barındırdığı için modeli iyileştirmek adına daha fazla seçenek tanımaktadır. Şekil 3.4 ile XGB algoritmasının çalışma prensibinde bulunan ağaçların dallanma şekli gösterilmiştir. Bu çalışma şekline seviye odaklı ağaç büyüme (level-wise tree growth) adı verilmektedir.

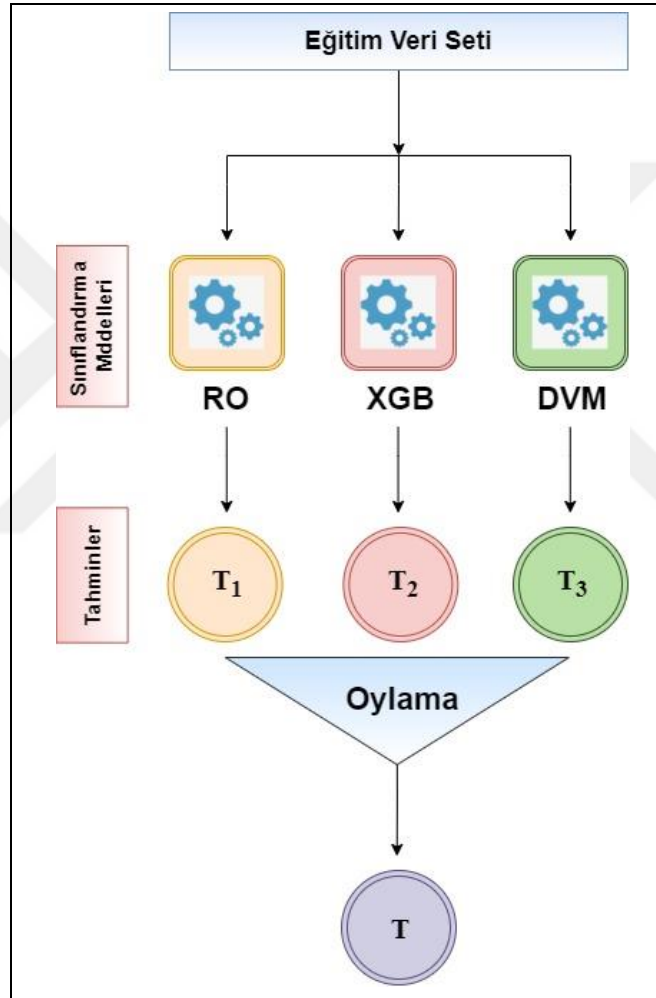


Şekil 3.4: Seviye odaklı ağaç büyümesi.

3.2.5. Oylama Tekniği

Oylama teknikleri birden fazla sınıflandırma algoritmasının veya aynı türde farklı parametrelere sahip sınıflandırma algoritmasının birleştirilerek ortaya daha yüksek performans sergilemektedir. Yöntem bir birleştirme tekniği olmasına rağmen farklı sınıflandırıcıların birlikte değerlendirilmesiyle yaygın bir kullanım alanına sahiptir [Doğaner, 2020]. Çoğunluk oylama yönteminde kullanılan her sınıflandırıcının tahminleri toplanarak en çok oyu alan sınıf belirlenir [Géron, 2019]. Ağırlıklı oylama yönteminde ise belirlenen olasılıklara göre sınıf ataması yapılmakta ve sınıflandırıcılara bir ağırlık atanmaktadır. Farklı sınıflandırıcılar aynı veri setine uygulanabildikleri için bir değerlendirme zenginliği oluşur ve doğruluk açısından sonucu artırmaktadır. Bu nedenle çeşitliliğin sağlanması durumunda, oylama topluluk öğrenmesi yönteminde yüksek performanslı tahminler elde edilmektedir [Polikar, 2012]. Oylama topluluk öğrenme yönteminde aynı veri setinde farklı türdeki sınıflandırıcıların kullanılması esaslı söz konusu olduğu için hızlandırma (boosting) ve torbalama (bagging) topluluk yöntemlerinden farklılık göstermektedir. Çünkü hızlandırma ve torbalama yöntemi aynı sınıflandırıcıdan çok fazla kullanarak veri

setinin farklı alt kümelerinin eğitilmesi durumu mevcuttur. Ayrıca oylama topluluk öğrenme yönteminde veri setinin tamamı işleme alındığı için üstün performans sağlayan bir etken olmaktadır. Şekil 3.5'te RO, XGB ve DVM algoritmaları kullanılarak bir oylama yönteminin çalışma şeması gösterilmiştir. Uzaktan algılamada görüntü sınıflandırma işlemleri için düşünülürse, üç farklı sınıflandırıcı değerlendirilerek sınıflandırma sürecinde pikselin ilgili arazi sınıfına atanması söz konusu sınıflandırma algoritmalarının değerlendirilmesiyle olur.



Şekil 3.5: Oylama yöntemi.

3.3. Sınıflandırma Sonuçlarının Analizi

Mühendislik alanında üretilen her sonucun güvenilirliğinin belirlenmesi veya ölçülebilmesi temel bir işlemdir. Uzaktan algılamada da sınıflandırma işlemi

sonucunda elde edilen tematik haritaların başarısı çeşitli değerlendirmelere tabi tutulur. Bu değerlendirmenin amacı uzaktan algılanmış görüntünün piksellerinin doğru arazi örtüsü sınıfına ne kadar etkili bir şekilde atandığı sayısal olarak görmektir [Rwanga and Ndambuki, 2017]. Hatanın az olması elde edilen doğruluk değerinin artmasını sağlamaktadır. Sınıflandırma sonucunda doğruluk değerlendirmesinin yapılmasıyla hataya neden olan durumlar anlaşılabilir ve hangi sınıflandırma yönteminin daha üstün performans sergilediği belirlenebilmektedir. İstatistiksel olarak bağımsız bir değerlendirme yapılabilmesi için eğitim verisinden bağımsız olarak test verileri doğruluk değerlendirilmesi kullanılmaktadır.

		j= Sütunlar (Referans)			Satır Toplamı ↓
		1	2	k	n_{i+}
i= Satırlar (Sınıflandırma)	1	n_{11}	n_{12}	n_{1k}	n_{1+}
	2	n_{21}	n_{22}	n_{2k}	n_{2+}
	k	n_{k1}	n_{k2}	n_{kk}	n_{k+}
Sütun Toplamı →		n_{+j}	n_{+2}	n_{+k}	n

Şekil 3.6: Hata matrisi gösterimi.

Şekil 3.6’de verildiği gibi sınıflandırma doğruluğunun belirlenebilmesi amacıyla hata matrisi kullanılmaktadır. Hata matrisi sınıflandırma sonucunda piksellerin hangi arazi örtüsüne atandığını belirten bir kare matristir. Şekilde $(k+1)$. sütun satırların toplamını ve $(k+1)$. satır ise sütunların toplamını temsil etmektedir. Toplam örneklem sayısı n ile gösterilmiştir. n_{ij} referans verisinde j grubundayken sınıflandırılmış görüntüde i sınıfına atanmış örnek sayısını ifade etmektedir. Sınıflandırma sonucunda i kategorisine atanmış sınıflandırma verisinin örneklem sayısı Eşitlik 3.2 ile, j kategorisi için ise referans verisindeki örneklem sayısı Eşitlik 3.3 ile gösterilmiştir.

$$n_{i+} = \sum_{j=1}^k n_{ij} \quad (3.2)$$

$$n_{+j} = \sum_{i=1}^k n_{ij} \quad (3.3)$$

Sınıflandırma işlemlerinde yaygın olarak kullanılan genel doğruluk, doğru olarak belirlenen piksel sayısının tüm piksellere oranı esas alınarak hesaplanır.

$$\text{Genel Doğruluk} = \frac{\sum_{i=1}^k n_{ii}}{n} \quad (3.4)$$

Genel doğruluk her bir sınıfa ait doğruluk performansı hakkında bilgi vermediği için üretici ve kullanıcı doğruluğu kavramları ortaya çıkmıştır. Üretici doğruluğu (ÜD) sınıfa atanmış piksellerin ilgili sınıfa ait referans verisi içindeki oranı ile ifade edilir ve 3.5'te formülü gösterilmiştir.

$$\text{Üretici Doğruluğu (ÜD)} = \frac{n_{jj}}{n_{+j}} \quad (3.5)$$

Bir sınıfa atanmış piksellerin gerçekte bu sınıfa ait olma doğruluğuna ise kullanıcı doğruluğu (KD) denir ve eşitlik 3.6'da nasıl hesaplandığı gösterilmiştir.

$$\text{Kullanıcı Doğruluğu (KD)} = \frac{n_{ii}}{n_{i+}} \quad (3.6)$$

Diğer doğruluk ölçüsü F-ölçütü (F-score), bir sınıfın sınıflandırılma başarısının belirlenmesi için kullanılır. Çok sınıflı sınıflandırmada, her bir sınıf için F-ölçütü hesaplanarak doğruluk değerlendirilmesi yapılabilir. Eşitlik 3.7'de belirtildiği üzere ilgili sınıfa ait üretici ve kullanıcı doğruluklar esas alınarak hesaplanmaktadır.

$$\text{F-ölçütü} = 2 \times \frac{\text{ÜD} \times \text{KD}}{\text{ÜD} + \text{KD}} \quad (3.7)$$

3.3.1. McNemar's Testi

Görüntülerin sınıflandırılmasında kullanılan algoritmaların performanslarına bağlı olarak elde edilen doğruluklar birbirlerinden farklılık göstermektedir. Ayrıca çeşitli bant kombinasyonlarının kullanılmasıyla elde edilen sınıflandırma doğrulukları aynı olmayacaktır. Hesaplanan genel doğrulukların değerlendirilip aralarındaki farkların istatistiksel olarak test edilmesi için Ki-kare dağılımının esas alındığı McNemar's (χ^2) testinin kullanılması söz konusudur. Bu test 2x2 boyutlu bir matris ile hesaplanan parametrik olmayan bir testtir [Foody, 2004]. Ki-kare tablo dağılımının esas alındığı değerlendirmede %90 veya %95 güven aralığı için tablodan bulunan değer hesaplanan değerden küçük olması halinde iki doğruluğun arasındaki farkın istatistiksel olarak anlamlı oldukları sonucuna varılır. Başka bir deyişle iki sınıflandırma sonucunun birbirlerinden istatistiksel olarak farklı olduklarını göstermektedir. Ki-kare tablosunun %95 güven aralığında ki değeri 3,84'dür. Eşitlik 3.9'da McNemar's (χ^2) testinin nasıl hesaplandığı gösterilmiştir. Formülde 2x2 boyutunda bir hata matrisi için i sınıflandırma algoritması tarafından hatalı j sınıflandırma algoritması tarafından doğru bulunan piksel sayısı n_{ij} ile ifade edilirken, n_{ji} j sınıflandırma algoritması tarafından belirlenen hatalı pikselleri ve i sınıflandırma algoritması tarafından belirlenen doğru pikselleri temsil etmektedir.

$$\text{McNemar's } (\chi^2) = \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} \quad (3.9)$$

3.4. Ters Mesafe Ağırlıklı Enterpolasyon Yöntemi (IDW)

Ters mesafe ağırlıklı enterpolasyon yöntemi sağladığı hesap ve çözüm kolaylığı nedeniyle mühendislik ve coğrafi bilgi sistemleri uygulamalarında yaygın olarak kullanılan bir yöntemdir [Göğsu ve Hastaoğlu, 2019]. Ters mesafe ağırlıklı enterpolasyon yöntemi, örnek noktaların doğrusal ağırlıklı bir kombinasyon kümesini kullanarak piksel değerlerinin belirlenmesi işlemidir (Childs, 2004). Bu yöntemde hesaplanacak piksel değeri, çevresinde bulunan komşu piksellerin uzaklığı ve büyüklüğünün bir fonksiyonudur. Ayrıca ters mesafe ağırlıklı enterpolasyon yöntemiyle belirlenen sınırdaki komşu piksellerden tahminler üretildiği için lokal bir

kestirim yöntemidir. Komşu piksellerin, tahmini yapılacak piksel üzerindeki etkisi mesafedeki değişime bağlıdır. Başka bir ifadeyle, ters mesafe ağırlıklı enterpolasyon yönteminde dayanak piksellere olan mesafe arttıkça uzakta olan piksellerin tahmin edilecek piksel değerine etkisi daha az olmaktadır. Ters mesafe ağırlıklı enterpolasyon işlemi Eşitlik 3.10'da gösterilmiştir. Eşitlikte u_0 tahmin edilen konumu, u_i ölçü konumunu, $\hat{Z}(u_0)$ N ölçümde ağırlıklandırılmış ortalamayı ve w_i ölçülen verilerin her birinin ağırlığı veya etkisini ifade etmektedir.

$$\hat{Z}(u_0) = \frac{\sum_{i=1}^N Z(u_i)w_i}{\sum_{i=1}^N w_i} \quad (3.10)$$

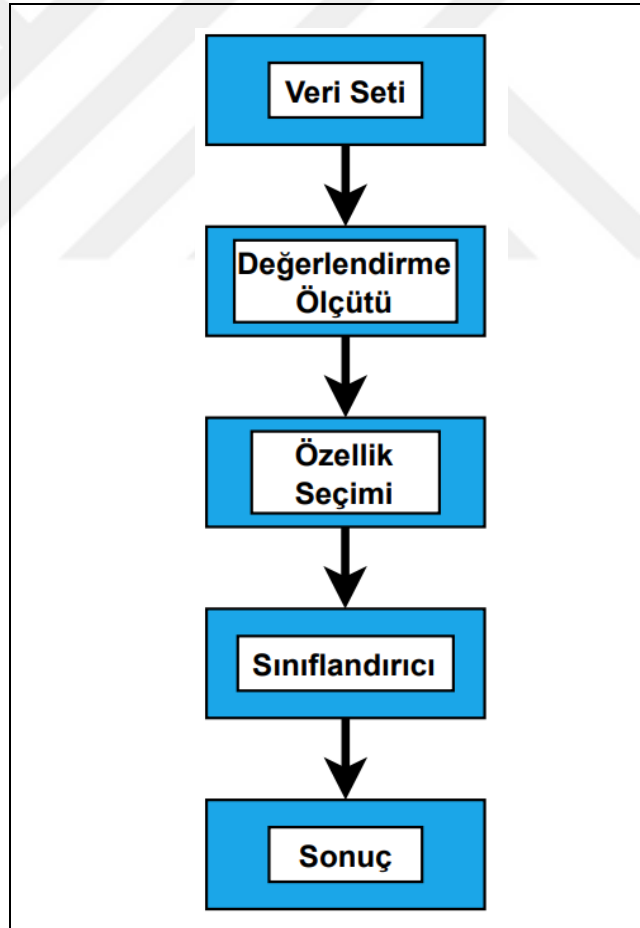


4. ÖZELLİK SEÇİMİ

Uzaktan algılamada arazi örtüsü/arazi kullanımı haritaları üretmek için sınıflandırma işlemine esas veri seti olarak uydu görüntülerinin kullanılmasının yanı sıra yardımcı veri setleri ile elde edilen çok boyutlu görüntülerde kullanılmaktadır. Bununla birlikte yüksek boyutlu veriler problem ile ilgisi olmayan, yanıltıcı; veya aralarında korelasyon olan özelliklerin bir arada oldukları veri setleridir. Veri sayının artmasıyla ortaya çıkan boyutsallık sorunu önemli bir sorundur [Hughes, 1968]. Bu tip verilere ön işlem adımı olarak boyut azaltmanın en basit yollarından biri olan özellik seçim işlemi uygulanmaktadır [Khalid et al., 2014]. Özellik seçimi işleminde belirlenen en iyi alt küme, öğrenme doğruluğuna en çok katkıda bulunan ve en küçük boyuta sahip olan kümedir [Ladha and Deepa, 2011]. Bant sayısının artmasıyla eğitim veri seti arasında gerçekleşen bu yakın ilişki ve sınıflandırma işlemi için gereken zaman maliyeti, boyut azaltma yöntemi ile kontrol altına alınabilmektedir. Bu sebeple sınıflandırma işlemi için tüm verinin yani veri setinde bulunan tüm bantların kullanılmasından ziyade, aralarında düşük korelasyon bulunan ve yararlı bilgiler sağlayan bantların seçilmesi gereklidir. Ayrıca özellik seçimi ile gereksiz bilgiler ortadan kaldırılarak gereksiz işlem yükünün önüne geçilerek zaman maliyeti düşürülmektedir. Burada zor olan kısım ise özellik seçimi ile alt kümelerin tanımlanması kısmıdır [Anthony and Ruther, 2007]. Uzaktan algılama çalışmaları için çok boyutlu verilere uygulanan özellik seçimi yaklaşımları ile özellik fazlalığının etkili bir şekilde azaltıp, sınıflandırmanın verimliliğini ve doğruluğunu artırmak oldukça önemlidir [Zhou et al., 2018]. Kısaca özellik seçim yöntemlerinin kullanılmasında üç ana avantajın olduğu söylenilebilir. Bunlardan ilki; bilgilendirici ve özelliklerin arasında düşük korelasyonlu bir kümenin seçilerek sınıflandırıcı performansının artırılmasıdır. İkincisi ise; daha az sayıda özellik kullanıldığı için işlem süresinin azalması ve son olarak veri boyutu ile örnek kümesinin boyutu arasındaki doğrudan ilişki nedeniyle eğitim örneklerinin sayısının sınırlı olduğu durumlarda kullanışlı olmasıdır [Kavzoglu, 2001]. Literatürde özellik seçimi amacıyla kullanılan çok fazla özellik seçimi yöntemi bulunmaktadır. Bu yöntemler genel olarak filtre tabanlı (filter), sarmalama (wrapper) ve ikisinin birleşimi olan hibirt (embedded) özellik seçimleri olmak üzere üç grupta incelenmektedirler [Kiala et al., 2019].

4.1. Filtre Tabanlı Yöntemler

Filtre tabanlı özellik seçimi sınıflandırıcı olmaksızın tamamen eğitim verilerine dayalı performans gösteren özellik seçim yöntemleridir [Kavzoglu and Mather, 2002]. Başka bir deyişle özellik alanında yapılan arama sınıflandırma sürecinden önce gerçekleştirilmektedir. Sınıflandırıcıdan bağımsız olarak hesaplanabildikleri için diğer özellik seçimi yöntemlerine göre daha az işlem ve bellek kaynağına ihtiyaç duyulmaktadır. Filtreleme tabanlı özellik seçimi yaklaşımları ile eğitim verileri arasında bulunan uzaklık, bağımlılık, tutarlılık ve bilgi gibi istatistiksel ölçüler gerçekleştirilmektedir [Budak, 2018]. Şekil 4.1’de belirtildiği gibi filtre tabanlı özellik seçim yönteminde önemli olan aşama kullanılan değerlendirme ölçütü yani istatistiksel olarak hangi yöntemin kullanıldığıdır.



4.1: Filtre tabanlı özellik seçim yöntemi.

Bu yaklaşım ile her bir banda ait olan önem derecesi bulunarak hangi bantların sınıflandırma işlemine dahil edileceği kararı verilir. Fakat bu özellik seçim yöntemi çok boyutlu görüntülere başarısız olabilmekle beraber sadece her banda ait bir puan belirlenmesi söz konusudur [Li et al., 2011]. Çalışmalarda kullanılan çeşitli filtre özellik seçimi yöntemleri mevcuttur, bu tez çalışması kapsamında Ki-kare testi, Kazanç Oranı (InfG) ve ReliefF algoritmaları verilere uygulanmıştır.

4.1.1. Ki-kare Testi

Bir filtre tabanlı özellik seçim tekniği olan Ki-kare özellik seçimi ki-kare (χ^2) istatistiğini temel alır ve bu yöntemde her bir özellik sınıf etiketine göre birbirlerinden bağımsız olarak değerlendirilir [Plackett 1983],[Mingers, 1989]. Spektral bantlarda temsil edilen arazi sınıflarına ait değerlerin dağılımı istatistiksel olarak incelenir. Hesaplanan istatistik değeri ne kadar büyük olursa değerlendirilen spektral bandın sınıflara ilişkin o oranda yararlı bilgi barındırdığı anlamına gelmektedir. Bu yöntemin temel çalışma esası χ^2 istatistiğinin gerçek değer ile beklenen değerden ne kadar farklı olduğunun hesaplanmasıdır. Eşitlik 4.1’de söz konusu istatistik formülünün nasıl hesaplandığı ifade edilmiştir. Eşitlikte r bir bantta bulunan farklı değerlerin sayısını, n sınıf sayısını temsil ederken, j sınıfındaki i değerine sahip piksellerin sayısı $O_{i,j}$ olarak ve j sınıfındaki i değerine sahip örneklerin beklenen sayısı $E_{i,j}$ olarak temsil edilmektedir. Elde edilen χ^2 istatistik değeri piksel ve sınıfların dağılımları hakkında bilgi vermektedir. χ^2 istatistik değerinin yüksek olması söz konusu değerlendirilmeye tabi tutulan bandın sınıflar ile yüksek bir ilişkiye sahip olduğu anlamına gelmektedir.

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^n \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \quad (4.1)$$

4.1.2. ReliefF

Kononenko [1994] tarafından iki sınıflı özelliklerin çözümünde kullanılan Relief yöntemine dayanarak ReliefF algoritması geliştirilmiştir. ReliefF algoritması sınıflara ilişkin özelliklere daha yüksek ağırlıklar atayan, konuyla ilgi seviyesi az olan

özellikleri hızlı bir şekilde ortadan kaldıran ve çoklu sınıflandırma problemlerinin çözümünde yüksek verimlilik sağlayan bir özellik seçim algoritmasıdır [Zhou et al., 2013]. Söz konusu algoritmanın uzaktan algılama uygulamalarında ise, spektral bantta bulunan ve birbirleri ile benzerlik gösteren piksel değerlerinin ayırt edilebilme yeteneğinden yararlanılarak kullanılmaktadır. K-en yakın komşuluk algoritmasının çalışma ilkesine benzerdir ve belirli bir mesafeye yakın uzaklıkların aynı gruba dahil olması esası söz konusudur. Algoritmanın formülü Eşitlik 4.2’de gösterilmiştir. Özellik değerlendirilmesindeki rastgeleliğin azaltılması ve nihai ağırlık olarak belirlenen ortalama değerin elde edilmesi için tüm süreç m kez tekrarlanmalıdır. Eşitlikte t özelliğinin ağırlığı ω_t , iki örnek arasındaki özellik uzaklık $diff_i()$, sınıf olasılığı $p(x)$, x örneğinin farklı bir sınıf örneği $H_j(x)$, en yakın aynı sınıf örneği $M_j(x)$, iterasyon sayısı m ve en yakın komşuların sayısı k ile temsil edilmektedir.

$$\omega_t^i = \omega_t^{i-1} + \sum_{c \neq class(x)} \frac{p(x)}{1 - p(class(x))} \sum_{j=1}^k diff_f(x, M_j(x)) / (m \times k) - \sum_{j=1}^k diff_f(x, H_j(x)) / (m \times k) \quad (4.2)$$

4.1.3. Bilgi Kazancı

Bilgi kazancı (InfG) makine öğrenmesi alanında yaygın bir şekilde kullanılan, entropi tabanlı bir özellik seçim metodudur [Li, 2012]. Entropi kavramı belirsizliğin bir ölçütü olarak karşımıza çıkmaktadır. X özelliği ve Y sınıf için bilgi kazancının hesaplandığı bir durum düşünülürse, X bilindiğinde Y değerindeki belirsizliğin azaltılmasına bilgi kazancı denir. Eşitlik 4.3’te verilen $H(Y)$ Y ’nin entropisi, $H(Y/X)$ X ’e dayalı Y ’nin şartlı entropisini temsil etmektedir. Bilgi kazancı oranı özellik seçimi yöntemi ile bir özellik kullanılarak bir sınıfa ait ne kadar fazla bilgi elde edildiği incelenir [Çölkesen, 2015]. Tekrarlı olarak özelliklere verilen puanlara göre, en yüksek puana sahip özellik (bant) en fazla bilgiyi içeren özelliktir. Eşitlik 4.3’te X özelliği için hesaplanan bilgi kazancı verilmiştir. Bu yöntem simetrik bir ölçüt olup X incelendikten sonra Y için kazanılan bilgi ve Y incelendikten sonra X için kazanılmış olan bilgi ile aynıdır.

$$InfG(X) = H(Y) - H(Y/X) \quad (4.3)$$

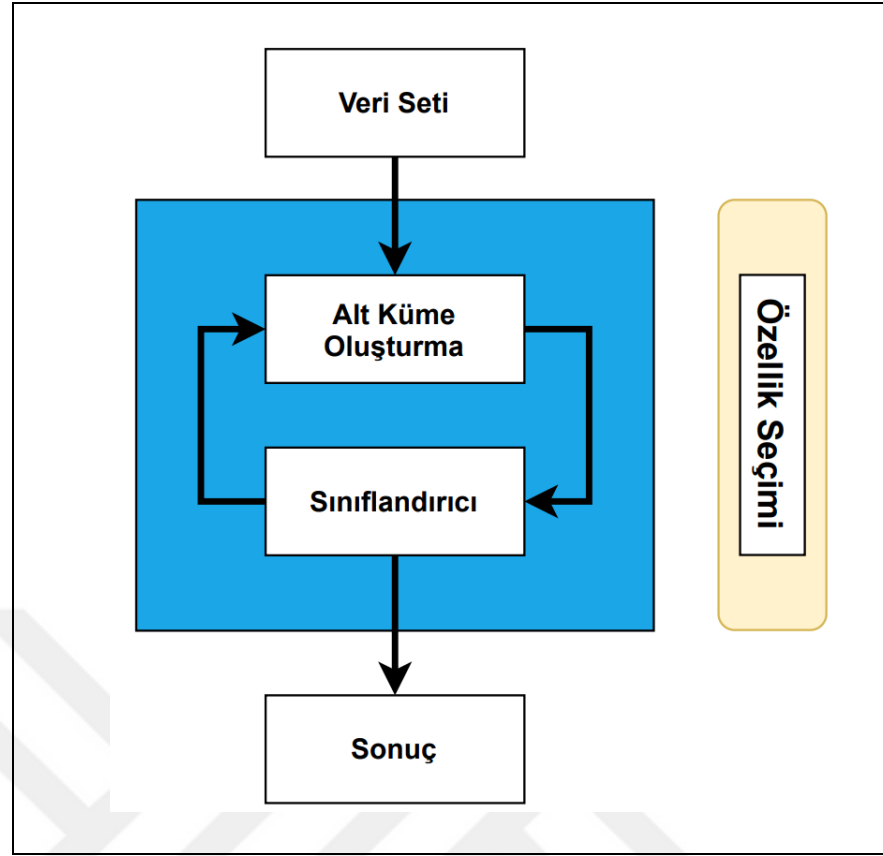
Eşitlik 4.3'te $Y = \{y_1, \dots, y_n\}$ ve $X = \{x_1, \dots, x_k\}$ olarak ele alındığında, $H(Y)$ eşitlik 4.4 ile hesaplanır. $H(Y/X)$ ise eşitlik 4.5 ile hesaplanmaktadır.

$$H(Y) = - \sum_{i=1}^n P(Y = y_i) \log_2 P(Y = y_i) \quad (4.4)$$

$$H(Y/X) = - \sum_{j=1}^l P(X = x_j) H(Y/X = x_j) \quad (4.5)$$

4.2. Sarmal Tabanlı Yaklaşımlar

Sarmalama tabanlı yaklaşımda, arama tekniklerinden herhangi biri kullanılarak alt küme oluşturulur ve bu alt kümeler öğrenme algoritmaları ile sınıflandırma doğruluğu açısından değerlendirilir [Dash and Liu, 1997]. Sarmalama tabanlı yaklaşım ile özellik seçiminde iki adım söz konusudur. Bu adımlardan ilki, öncelikle özelliklerin bir alt kümesinin aranıp, daha sonra ise seçilen özellikler en yüksek öğrenme performansına ulaşılan kadar tekrarlı olarak değerlendirilir [Kiala et al., 2019]. Şekil 4.2'de çalışma şekli gösterilen sarmal özellik seçiminin sürecinde, en yüksek öğrenme performansı elde edilene kadar veya istenen sayıda seçilen özellik seçilene kadar yinelemeli olarak çalışır. Bu değerlendirme kriterleri yapılan çalışmaya veya işlemi gerçekleştiren analiste göre değişebilmektedir. Bu kriterlerden biri doğruluktur ve seçilen alt kümelerin sınıflandırma doğruluk performanslarına göre sarmal yaklaşımların uygulanması mümkündür.



4.2: Sarmal özellik seçim yaklaşımı.

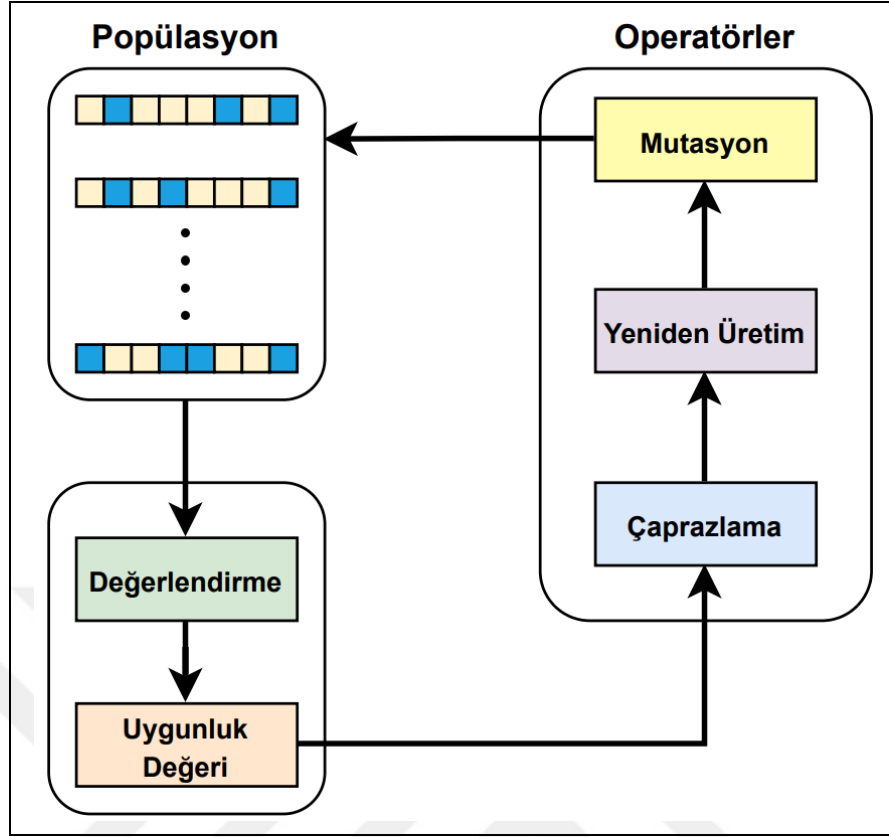
Sarmalama özellik seçimi yönteminin olumsuz tarafı d özelliğe sahip bir veri için 2^d arama alanına sahip olmasıdır ve özellik sayısı fazla olduğunda pratik bir çözüm olmamasıdır [Li et al., 2017]. Başka bir deyişle sarmal özellik seçim yöntemlerinin uygulanması filtre özellik seçim yöntemlerine göre zaman açısından maliyetlidir. İleri yönde seçim (AİS), genetik algoritma (GA) ve öz yinelemeli özellik seçimi (ÖÖS) sarmal tabanlı özellik seçim teknikleridir. Sezgisel yaklaşımlar olarak da adlandırılan söz konusu sarmal teknikler bir uygunluk fonksiyonu (fitness function) ile beraber uygulanmaktadır. Kavzoglu [2001] tarafından yapılan doktora çalışmasında diverjans, dönüştürülmüş diverjans, Bhattacharyya uzaklığı, Hotelling's T^2 uzaklığı, Wilks' $\tilde{E}$ testi gibi birçok istatistiksel tabanlı uygunluk fonksiyonunu, uzaktan algılanmış görüntüye sınıflandırma işleminde yararlanılacak spektral bantların belirlenmesi için sarmal tabanlı özellik seçim işlemlerinde kullanılmıştır. Pal [2006] tarafından yapılan çalışmada GA ve ÖÖS özellik seçim işlemlerinde DVM makine öğrenme algoritmasını uygunluk fonksiyonu olarak değerlendirmiştir.

Tez çalışması kapsamında RO algoritması uygunluk fonksiyonu olarak seçilmiş ve bu üç yöntem ile birlikte genel doğrulukları açısından değerlendirilerek, özellik

seme iřlemi iin kullanılmıřtır. zellik seim iřlemlerinin nesne bazlı sınıflandırma iřlemlerindeki etkisini gzlemlemek iin yapılan bir alıřmada eřitli algoritmaların yanı sıra S'ye dayalı RO ve XGB kullanılmıřtır [Georganas et al., 2018].

4.2.1. Genetik Algoritma

Genetik algoritma bir nesildeki en uygun bireylerin hayatta kalma ve yeni nesli retme olasılıklarının daha yksek olduėu varsayımsal doėal seleksiyon mekanizmasından esinlenerek geliřtirilmiřtir [Kavzoglu, 2001]. Uyarlanabilir bir arama tekniėi olan genetik algoritma yaklařımı incelenen sorunun zelliėine gre birden fazla maksimumun veya minimumun bulunması iin kullanılır [Goldberg, 1988]. Ayrıca tm olası zmlerin deėerlendirilmesi hesaplama sresi aısından maliyetli olduėu durumlarda en uygun zmlerin aranması iin kullanılır. Bu yntem rastgele seilmiř bireylerden oluřan ilk poplasyon ile bařlar ve poplasyonda bulunan her bir birey zlmesi amalanan problem iin potansiyel oluřturur. Bireyler tekrarlı iřlemler ile geliřerek tekrarlamaların sonucunda yeni nesiller ortaya ıkmaktadır. Nesillerde bulunan poplasyondaki her birey iin eřitli uygunluk ltleri kullanarak deėerlendirme yapılarak genetik operatrler (yeniden retim, arpazlama ve mutasyon) aracılıėıyla bir sonraki nesil poplasyonu oluřturulmuř olur. Algoritmada sonlandırma řartları saėlanan kadar sre devam eder. Genetik operatrlerden yeniden retim (reproduction) var olan neslin yerine yeniden nesil oluřturmak veya nesildeki birkaç bireyin yenileri ile yer deėiřtirmesi olarak tanımlanmaktadır. Diėer operatr olan arpazlama (crossover) en nemli operatrdr ve poplasyonda bulunan iki farklı bireyin yapılarının rastgele bir biimde birleřtirilmesi ile yeni bireylerin oluřturulması iřlemine verilen isimdir [İři ve Korukoėlu, 2003]. Mutasyon (mutation) ise bir bireyin genlerinin bir ya da bir kaının yerlerinin deėiřtirilmesiyle gerekleřtirilen iřlemdir. řekil 4.3'te genetik algoritma yaklařımının ařamaları gsterilmiřtir.



4.3: Genetik algoritma yaklaşımı.

4.2.2. Ardışık İleri Yönde Seçim Yaklaşımı

Whitney [1971] tarafından önerilen ardışık ileri yönde seçim (SFS) yönteminde boş bir küme ile işleme başlanır ve özellikler birer birer değerlendirilir. Değerlendirme işlemi sonucuna göre en iyi olan özellik tespit edilir ve daha sonra ikinci en iyi özelliğin bulunup alt kümeye eklenmesi ile istenilen sayıda özellik seçilene kadar yinelemeli olarak çalışan bir özellik seçim yöntemidir [Kavzoglu, 2001]. Söz konusu özellik seçim algoritmasının performansı kullanılan sınıflandırma algoritmasının başarı oranına bağlıdır. Yerleştirme etkisi olarak adlandırılan (nesting problem) sorundan dolayı uygun olmayan bir çözüm üretilir. Bu etkinin sebebi ise seçilen özelliklerin özellik kümesinde tutulması ve eklenen özellikler ile aralarında etkileşim olabileceği gerçeğinden kaynaklanmaktadır. Arama uzayında yinelemeli olarak işlem yapılmasının doğal bir sonucu olarak zaman maliyetini artmaktadır. Ardışık ileri seçim yöntemi özetle; en iyi özelliğin seçilmesiyle başlanarak yinelemeli olarak yapılan sınıflandırma sonuçlarına göre yeni özelliklerin eklenerek istenilen sayıda özelliğin

seçilmesi işlemidir. Uzaktan algılamada bu özellikler spektral bantlardır ve yüksek boyutlu uzaktan algılanmış verilere söz konusu yöntemle istenilen sayıda bandın seçilmesi için uygulanmaktadır.

4.2.3. Özyinelemeli Özellik Seçim Yaklaşımı

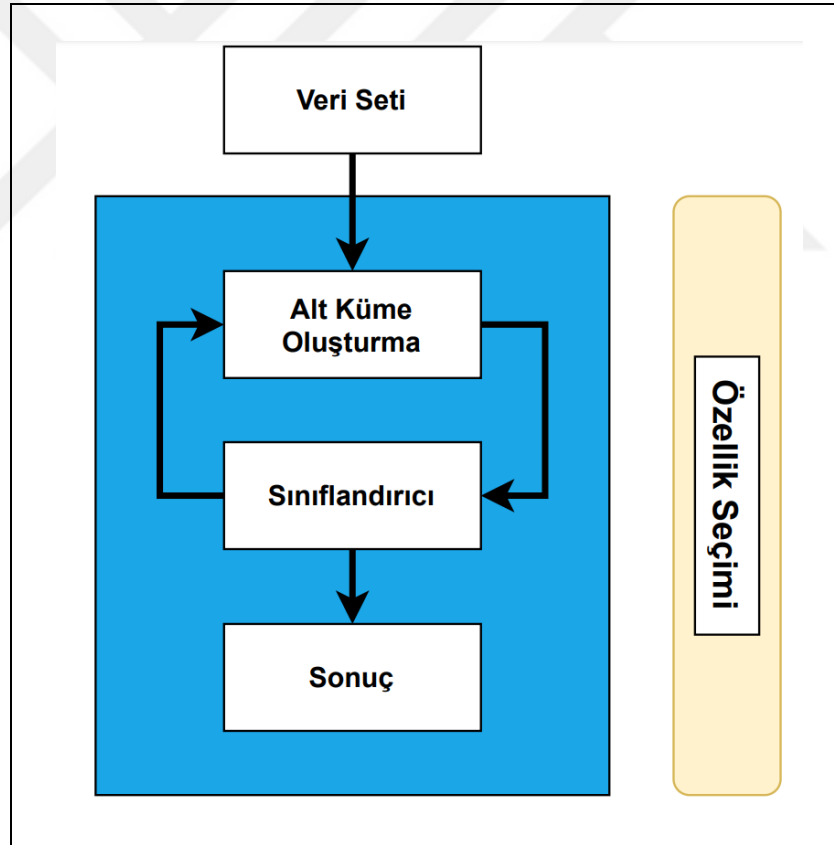
Özyinelemeli özellik seçim yöntemi özellik sıralama tekniğine dayanan açgözlü bir algoritmadır [Guyon et al., 2002]. Bu özellik seçim tekniği tüm öngörücüler kümesi üzerinde bir model oluşturarak ve her öngörücü için bir önem puanı hesaplayarak başlar. En az önemli olanlar kaldırılarak yeniden model oluşturulur ve önem puanı yeniden hesaplanarak istenilen sayıda en iyi özelliklerin belirlenmesi temel esastır. Son hesaplanan sınıflandırma doğruluğu azaltılmadan optimal istenilen sayıda özellik barındıran alt küme elde etmek için uygulanmaktadır. Ayrıca tespit edilmek istenen özellik kümesi özyinelemeli özellik seçimi için bir ayar parametresidir. Diğer önemli parametre ise özelliklerin önem derecesini belirlemek için kullanılacak olan öğrenme algoritmasıdır. Rastgele orman algoritmasının sınıflandırıcı olarak ele alındığı bir özyinelemeli özellik seçimi işleminde her turda yeni bir modelin öğrenilmesi, çapraz doğrulama ile doğruluk değerlendirmelerinin yapılması, işleme alınan her özellik için önem metriklerinin özelliğinin analiz edilmesi ve ardışık tur için kullanılacak özellik setininin değiştirilmesi söz konusudur. Daha sonra en zayıf performans gösterenler öğrenme sırasında model tarafından tahmin edilen önem değişken metriği kullanılarak tespit edilir [Demarchi et al., 2020].

4.3. Gömülü Yöntemler

Bir diğer boyut azaltma yaklaşımı olan gömülü özellik seçim yöntemi optimum özellik alt kümesinin seçilmesi sınıflandırıcı tarafından gerçekleştirilir. Öğrenme süreci tarafından yönlendirilen arama süreci nedeniyle yüksek doğruluk sağlamanın yanında sarmalama özellik seçim yönteminin aksine daha az hesaplama maliyetine ihtiyaç duyarlar çünkü özellik seçimi sınıflandırma modeli ile gerçekleştirilir [Moghimi et al., 2018]. Diğer bir ifadeyle tüm özellik alt kümeleri için sınıflandırma algoritmasını yinelemeli olarak çalıştırmaya gerek olmadığı anlamını taşımaktadır.

Şekil 4.4'te bir gömülü özellik seçim yaklaşımının çalışma prensibi ve sınıflandırma algoritması ile olan ilişkisi görülmektedir.

Sarmalama yöntemlerinde olduğu gibi, gömülü yöntemlerde de özellik seçimi işlemi kaçınılmaz olarak kullanılan sınıflandırma algoritmasına bağlıdır ve seçilen özellik alt kümesi diğer sınıflandırıcılar için etkili olmayabilir [Hira and Gillies, 2015]. Bunun temel nedeni gen kümesinin sınıflandırıcı oluşturulduğunda inşa edilmesi ve seçimin sınıflandırıcının kurduğu hipotezden etkilenmesidir. Ayrıca gömülü özellik seçim yöntemleri sınıflandırıcı modeline bağlı olarak gerçekleştirildiği için sınıflandırıcının zayıflıkları ile sınırlı olabilmektedirler [Georganos et al., 2018]. Gömülü özellik seçim teknikleri olan ve tez çalışmasında değerlendirilen algoritmalar, Hilbert-Schmidt Serbestlik Kriteri Lasso (HSIC Lasso), rastgele orman (RO) ve hafif gradyan artırma makineleri (LightGBM)'dir.



Şekil 4.4: Gömülü özellik seçim yöntemi.

4.3.1. LightGBM ile Özellik Seçimi

Microsoft tarafından 2017 yılında geliştirilen hafif gradyan artırma makineleri (LightGBM) karar ağaçlı tabanlı yeni nesil sınıflandırma yöntemlerinden birisidir. Yüksek boyutlu veriler ile çalışmanın bir dezavantajı olan zaman maliyetini düşürmek için önerilmiştir [Ke et al., 2017]. LightGBM algoritmasını diğer gradyan artırma yöntemlerinden ayıran en önemli özelliklerinden biri karar ağaçlarının dallanma şeklinin farklı olmasıdır. Bu yöntemin çalışma ilkesi Şekil 4.5'te gösterilmiştir. LightGBM yönteminde karar ağaçları yaprak odaklı (leaf-wise) büyüme gösterirken, diğerler artırma algoritmalarında seviye odaklı (level-wise) büyüme durumu söz konusudur.

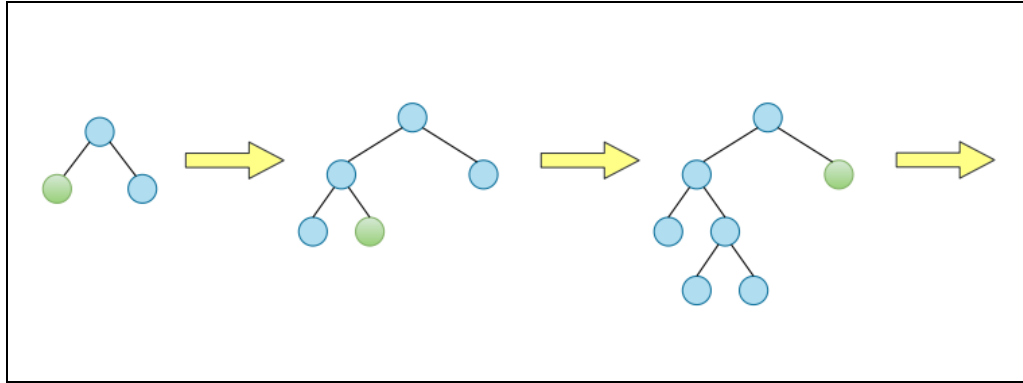
Bu yöntemde işlem hızının artmasını sağlayan gradyan alt yapılı tek yönlü örnekleme (Gradient-Based One-Side Sampling) ile ayrıcalıklı öznitelik destekleme (Exclusive Feature Bundling) algoritmaları mevcuttur [Zhang and Gong, 2020]. Gradyan tabanlı tek yönlü örneklemede karar ağaçlarına ait doğruluk oranı korunarak önemli veriler kullanılır. Bunun sonucunda veri dağılımı çok fazla etkilenmeden veri sayısında azaltma yoluna gidilmiş olur. Ayrıcalıklı öznitelik destekleme algoritmasında ise seyrek halde bulunan özelliklerin daha yoğun hale getirilerek işlem karmaşıklığının azaltılması esastır [Ustuner and Sanli, 2019].

Diğer gradyan artırma karar ağaçlarından farklı olarak, LightGBM'de varyasyon kazancının hesaplanma işlemi zayıf ve güçlü öğrenenler altında gerçekleşir ve her bir özelliğin eğitim verileri, tüm ağaçlara bölmek için kaç kez kullanıldığına bağlı olarak FIS_i önemlilik puanı hesaplanır [Tang et al., 2020]. Eşitlik 4.7'de gösterildiği şekilde hesaplanan formülde, ω_i her bir özelliğin ağırlığını, x_i özellik kümesini temsil etmektedir. Böylece özelliklere belirli bir önem sırası atanması mümkündür.

$$FIS_i = \{s | s = \omega_i x_i \} \quad (4.7)$$

LightGBM hızlı işlem yapabilme, daha büyük verileri analiz edebilme, daha iyi doğruluk sağlayabilme yeteneklerine sahiptir [Daoud, 2019]. Ayrıca daha az bilgisayar belleği ile çalışabilmesiyle diğer gradyan artırma yöntemlerinden ayrılmaktadır. Bu yöntemin büyük boyutlu verileri işlemek için daha az belleğe ihtiyaç duyması,

özellikle teknolojinin gelişmesine bağlı olarak uzaktan algılama alanında artan yüksek boyutlu verilerin analiz edilmesi sorununa alternatif bir çözüm yöntemi olmaktadır.



Şekil 4.5: Yaprak odaklı ağaç büyümesi.

4.3.2. Hilbert-Schmidt Serbestlik Kriteri Lasso ile Özellik Seçimi

Yamada et al. [2014] tarafından geliştirilen Hilbert-Schmidt Serbestlik Kriteri - Lasso (HSIC-Lasso) yöntemi doğrusal olmayan ve son yıllarda geliştirilmiş yeni bir özellik seçim yöntemidir. Hem özellik seçimi hem de tahmin algoritmasını içeren en az mutlak büzülme ve seçme operatörü (Lasso) yönteminden farklı olarak HSIC Lasso yalnızca özellik seçim işlemi için kullanılmaktadır [Takahashi et al., 2020]. Algoritma Hilbert-Schmidt Serbestlik Kriteri (HSIC) ve Lasso metodunun bir kombinasyonu olarak geliştirilmiştir. Lasso özellik seçiminden esinlenerek doğrusal olmayan girdi-çıkış bağımlılığına (input-output dependencies) sahip olacak şekilde, çekirdek yöntemine dayanan HSIC algoritmasının kullanılması mantığına göre tasarlanmıştır. Başka bir ifadeyle algoritma bir Lasso problemini çözümü ve çekirdek işlemi olan HSCI tekniğinin uygulanmasıyla gereksiz olmayan özelliklerin seçimine olanak sağlamaktadır [Afshar & Usefi, 2021]. Lasso doğrusal bir yöntemdir buna nedenle tahmin ediciler ve hedef arasında bağımlılığın doğrusal olmamasının beklendiği uygulamalarda kullanılması uygun olmayabilir [He et al., 2014]. Bunun üstesinden gelmek için doğrusal olmayan HSIC Lasso metodu geliştirilmiştir. Bu yöntemin avantajı ise, verimli bir optimal çözümün ortaya konulmasıdır. Böylece yüksek boyutlu verilere de özellik seçim işlemi için uygulanabilmektedir. Eşitlik 4.6 ile yöntemin ne şekilde formüle edildiği ifade edilmiştir.

$$\min_{\alpha \in \mathbb{R}^p} \frac{1}{2} \|\bar{L} - \sum_{k=1}^p \alpha_k \bar{K}^{(k)}\|_{Frob}^2 + \lambda \|\alpha\|_1, \quad (4.8)$$

$$\alpha_1, \dots, \alpha_p \geq 0$$

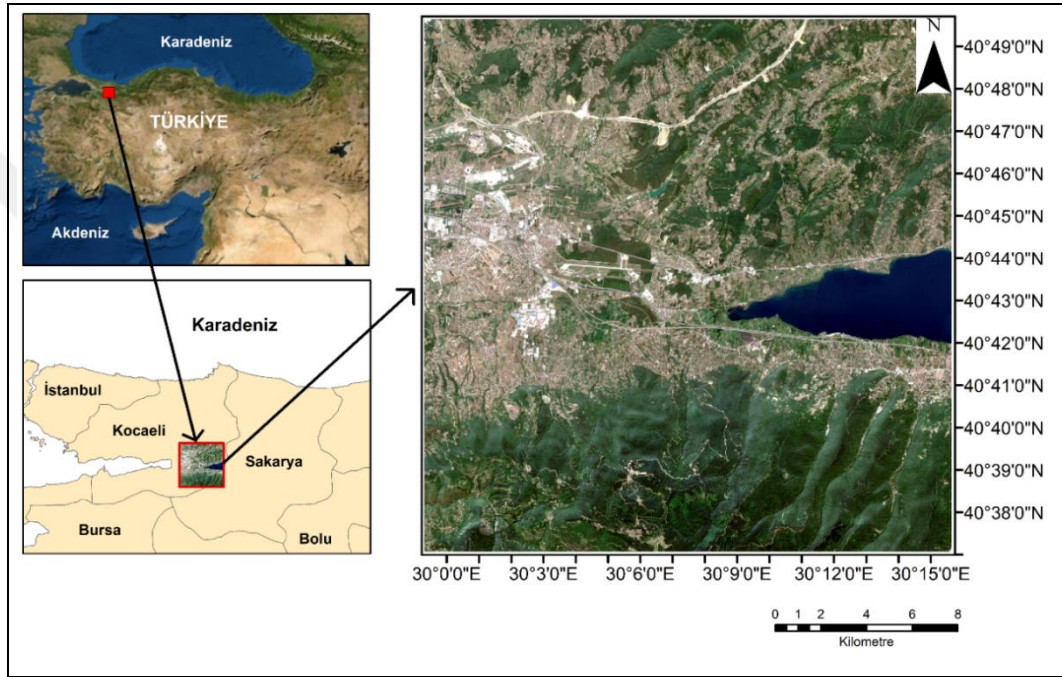
Eşitlikte $\|\cdot\|_{Frob}$ Frobenius normunu, $\|\cdot\|_1$ ise ℓ_1 normunu göstermektedir. Ayrıca α regresyon katsayı vektörünü, λ düzenleme parametresini, $\bar{L}$ çıktı çekirdek matrisini ve $\bar{K}^{(k)}$ girdi çekirdek matrisini temsil etmektedir.

4.3.3 Rastgele Orman Algoritması ile Özellik Seçimi

Rastgele orman (RO) çok miktarda karar ağacının kullanılmasıyla karar ağacı ormanı oluşturan bir sınıflandırma tekniğidir [Breiman, 2001]. Söz konusu algoritmanın içerdiği değişkenlerin önemlilik derecelerinin belirlenebilmesi özelliğiyle etkili bir özellik seçim yöntemi sunar [Genuer et al., 2010]. Ormanda bulunan karar ağaçlarının eğitilmesi için veri setinden rastgele bir şekilde alt kümeler üretilir. Rastgele olarak üretilen alt kümelerden 2/3'ü karar ağaç yapısını meydana getirmek üzere, kalan 1/3'lük kısmı ise (Out of Bag, OOB) ağaç yapısının test edilmesi için kullanılmaktadır. OOB verileri sınıflandırma sürecinde özelliklerin önemini belirleyerek özellik seçimi için kullanılmaktadır. Her değişkenin doğruluğu, özelliklere bir önem seviyesi atamak ve özelliklerin sıralı bir listesini oluşturmak amacıyla kullanılacak bir Z puanı elde etmek için standart hata ile normalleştirilir [Löw et al., 2013]. Buna bağlı olarak bir gömülü özellik seçim metodu olan RO, işleme alınan özellikler için belirlenen önemlilik derecesine göre bir sıralama gerçekleştirir.

5. ÇALIŞMA ALANI ve VERİ SETİ HAZIRLAMA

Tez kapsamında yardımcı verilerin oluşturulup çeşitli özellik seçim yöntemlerinin uygulanarak, sınıflandırma doğruluklarına olan katkılarının araştırılması amaçlanmıştır. Bu bağlamda Şekil 5.1’de gösterildiği üzere büyük bir bölümü Kocaeli il sınırı içinde olan yaklaşık 537 km²’lik yüz ölçüme sahip bir çalışma alanı belirlenmiştir.



Şekil 5.1: Çalışma alanı.

Çalışma alanının tespit edilmesinde, arazinin sahip olduğu yüksek topoğrafik değişimin yanı sıra farklı arazi sınıflarını içermesi etkili olmuştur. Arazideki mevcut topoğrafik farklılık, yardımcı veri olarak kullanılan ve sayısal yükseklik modelinden (SYM) üretilen yükseklik, eğim ve bakı haritalarının etkilerinin daha doğru bir şekilde değerlendirilmesi için önem arz etmektedir. Belirlenen yedi arazi sınıfı orman, şehir, yol, su, toprak, ekili tarım alanı ve ekili olmayan tarım alanlardır. Ayrıca her bir arazi sınıfı için 700 eğitim ve 300 test pikseli belirlenmiş olup eğitim için toplam 4900 piksel, test için ise 2100 piksel kullanılmıştır. Her sınıf için eşit sayıda eğitim ve test verilerinin kullanılmasıyla, sınıflandırma aşamasında meydana gelebilecek olası önyargının engellenmesi amaçlanmıştır.

Tez çalışmasında sınıflandırma işlemlerinin gerçekleştirilmesi ve tematik harita üretimi için temel uydu verisi olarak çalışma alanını kapsayan 08 Ağustos 2020 tarihinde kaydedilmiş Sentinel-2A uydu görüntüsü kullanılmıştır. Ayrıca yardımcı veri olarak kullanılması amacıyla ALOS-PALSAR uydusundan 12,5 metre mekânsal çözünürlüğe sahip olan SYM verisi elde edilmiştir. Bu nedenle sağlıklı bir değerlendirme yapabilmek adına, Sentinel-2A uydu görüntüsünün çalışmada kullanılan spektral bantları 12,5 metre mekânsal çözünürlüğe en yakın komşuluk yöntemiyle yeniden örneklendirilmiştir. Spektral bantlardan elde edilen yardımcı veriler yeniden örnekleme işleminden sonra üretilerek çalışmada kullanılmıştır. Tez çalışması kapsamında uydu görüntüsünün mavi, yeşil, kırmızı, kırmızı kenar-1, kırmızı kenar-2, kırmızı kenar-3, NIR, daraltılmış NIR, SWIR-1 ve SWIR-2 spektral bantları değerlendirilerek bütün işlemler gerçekleştirilmiştir.

Avrupa Komisyonu'nun yürüttüğü Copernicus programı ile geliştirilen Sentinel-2A uydusu 23 Haziran 2015 tarihinde uzaya fırlatılmış olup 10 günlük bir zamansal çözünürlüğe sahiptir. Uydunun farklı çözünürlüklerde 13 spektral bandı bulunmaktadır [Wang et al., 2019]. Tablo 5.1'de Sentinel-2A uydusuna ait genel özellikler verilmiştir.

Tablo 5.1: Sentinel-2A uydusu teknik özellikleri.

Fırlatılma tarihi ve yeri	Haziran 2015, Fransa
Yörünge Yüksekliği	786 km
Zamansal Çözünürlük	10 gün
Mekânsal Çözünürlük	10 m, 20 m ve 60 m
Radyometrik Çözünürlük	12 bit
Spektral Bant Sayısı	13

Tablo 5.2'de Sentinel-2A uydusuna ait görüntülerin spektral özellikleri belirtilmiştir. Tabloda görüldüğü üzere Sentinel-2A uydu görüntüleri 10 m, 20 m ve 60 m mekânsal çözünürlüğe sahiptir. Özellikle kullanıcılara 10 m mekânsal çözünürlükte, ücretsiz ve Sentinel-2B uydusuyla birlikte ele alındığında 5 günlük zamansal çözünürlükte veri sağlaması birçok çalışmada Sentinel uydu görüntülerinin tercih edilmesi söz konusu olmuştur.

Tablo 5.2: Sentinel-2A uydusu spektral özellikleri.

Bant No	Bant Adı	Dalga Boyu (nm)	Mekansal Çözünürlük (m)
1	Kıyı aerosol	433-453	60
2	Mavi	458-523	10
3	Yeşil	543-578	10
4	Kırmızı	650-680	10
5	Kırmızı Kenar-1	698-713	20
6	Kırmızı Kenar-2	733-748	20
7	Kırmızı Kenar-3	773-793	20
8	Yakın Kızılötesi	785-900	10
8A	Daraltılmış yakın kızılötesi	855-875	20
9	Su buharı	935-955	60
10	Kısa Dalga Kızılötesi	1365-1385	60
11	Kısa Dalga Kızılötesi-1	1565-1655	20
12	Kısa Dalga Kızılötesi-2	2100-2280	20

Özellik seçim ve sınıflandırma işlemlerinin tamamı Python 3.8.3 programlama dilinde, Anaconda'nın Jupyter Notebook ara yüzü kullanılarak analiz edilmiştir. Python programlama dili ile üretilen sonuç görüntülerin görselleştirilmesi ve analiz edilmesinin yanı sıra, yardımcı verilerin analizi, değerlendirilmesi ve işlenmesi için ArcGIS, ENVI, SNAP ve QGIS gibi çeşitli programlar kullanılmıştır. Modellerin oluşturulması, doğruluklarının hesaplanması ve veri normalizasyonu gibi işlemler için scikit-learn kütüphanesinden faydalanılmıştır. Pandas kütüphanesi eğitim ve test verilerinin okunması için numpy kütüphanesi ise eğitim ve test verilerinin dizi formatında değerlendirilebilmesi için kullanılmıştır. Ayrıca tez kapsamında kullanılan bilgisayar Intel(R) Core(TM) i5-9400 CPU @ 2.90GHz işlemciye, 16.00 GB RAM'e, 64 bit işletim sistemine ve Intel(R) UHD Graphics 630 ekran kartı özelliklerine sahiptir.

5.1. Spektral Bantlardan Üretilen Yardımcı Veriler

Tezin amacı kapsamında Sentinel-2A uydu görüntüsü ile arazi örtüsü/kullanımına uygun olarak daha yüksek sınıflandırma doğruluğunda tematik harita üretimi için çeşitli yardımcı veriler kullanılmıştır. Uydu görüntüsünün bantları kullanılarak elde edilen yardımcı veriler gri düzey eş oluşum matris formülleri kullanılarak hesaplanan doku özellikleri, temel bileşenler analizi ile tespit edilen bileşenler, ton-doygunluk-yoğunluk renk uzayı dönüşümü ile elde edilen özellikler, spektral indisler ve morfolojik filtrelerin uygulanmasıyla elde edilen özelliklerdir.

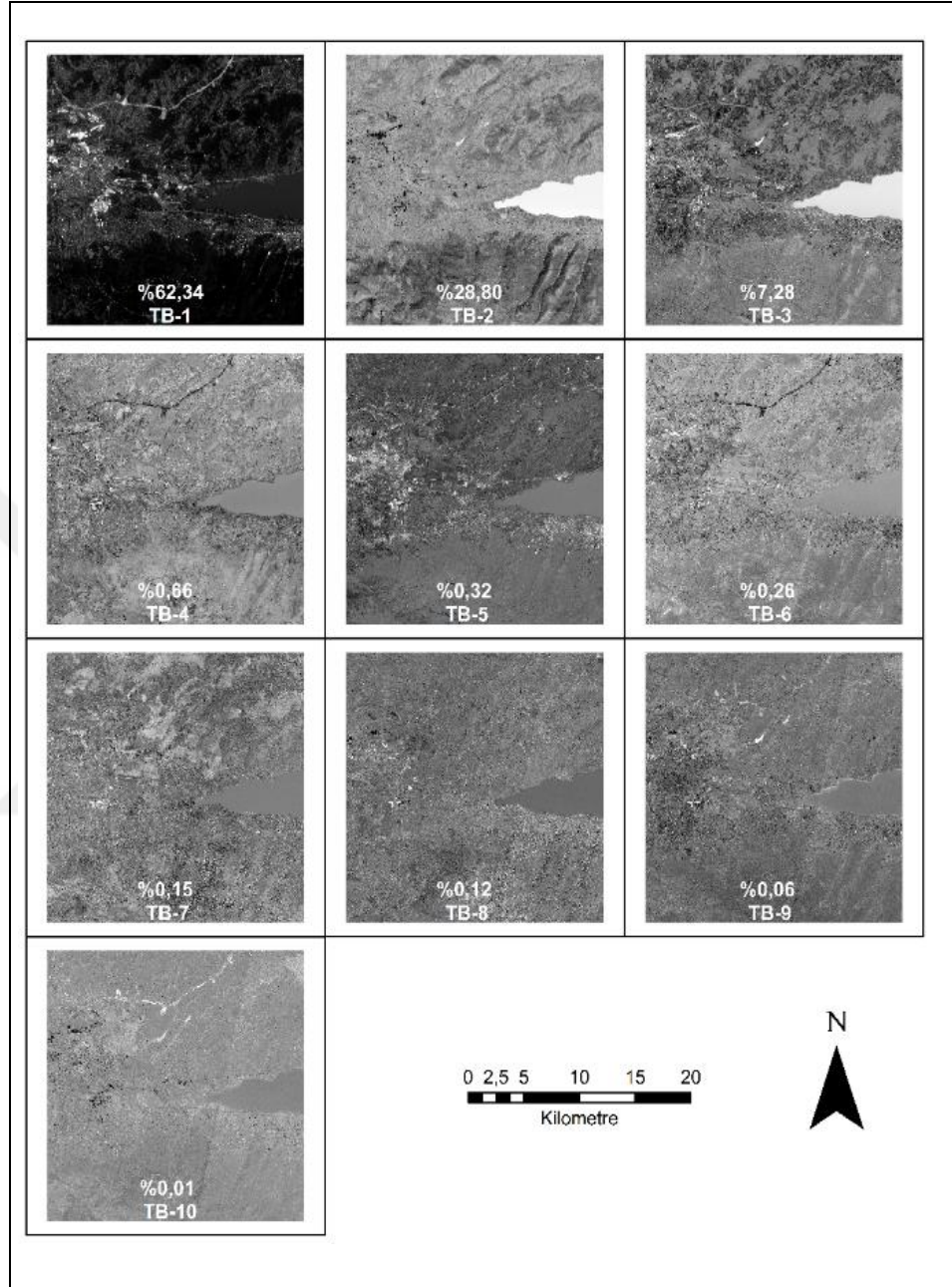
5.1.1. Gri Düzey Eş Oluşum Matrisleri ile Üretilen Doku Özellikleri

Görüntü sınıflandırma işlemine katkı sağlayıp sağlamadığını analiz etmek için üretilen ilk yardımcı veri seti, gri düzey eş oluşum matris eşitlikleri kullanılarak hesaplanan doku özellikleridir. Doku düzenli bir frekansta tekrarlayan desenlerin oluşturduğu kombinasyon olarak tanımlanmaktadır. Ayrıca görüntüde ilgi duyulan nesnelerin tanımlanabilmesi için önemli bir özelliktir. Tez çalışması kapsamında yardımcı veri seti olarak kullanılan doku özellikleri 5x5 pencere boyutu kullanılarak; 0°, 45°, 90° ve 135° olmak üzere dört farklı yönde hesaplanmıştır. Tez çalışması kapsamında Sentinel-2A uydu görüntüsünün 10 bandına daha önce Tablo 2.1’de gösterilen eşitliklerle hesaplanan açısıl ikinci moment, entropi, farklılık, homojenlik, korelasyon, ortalama, varyans ve zıtlık 8 doku özelliği belirtilen yönlerde görüntüye uygulanarak toplam 80 bant ek veri olarak değerlendirilmek üzere üretilmiştir

5.1.2. Temel Bileşenler Analizi

Veri setindeki çeşitliliği artırmak için temel bileşenler analizi ile elde edilen bileşenler yardımcı veri seti olarak değerlendirilmiştir. Uydu görüntülerine uygulanan temel bileşenler analizi ile üretilen yeni görüntünün yorumlanabilir olma özelliği orijinal görüntüden daha iyi olmaktadır. Uydu görüntüsüne uygulanan temel bileşen analizi ile çalışmada kullanılan Sentinel-2A görüntüsünün 10 bandına karşılık gelen

bileşenler hesaplanmıştır. Şekil 5.2’de hesaplanan temel bileşenlere ait varyans yüzdelik oranları gösterilmiştir.



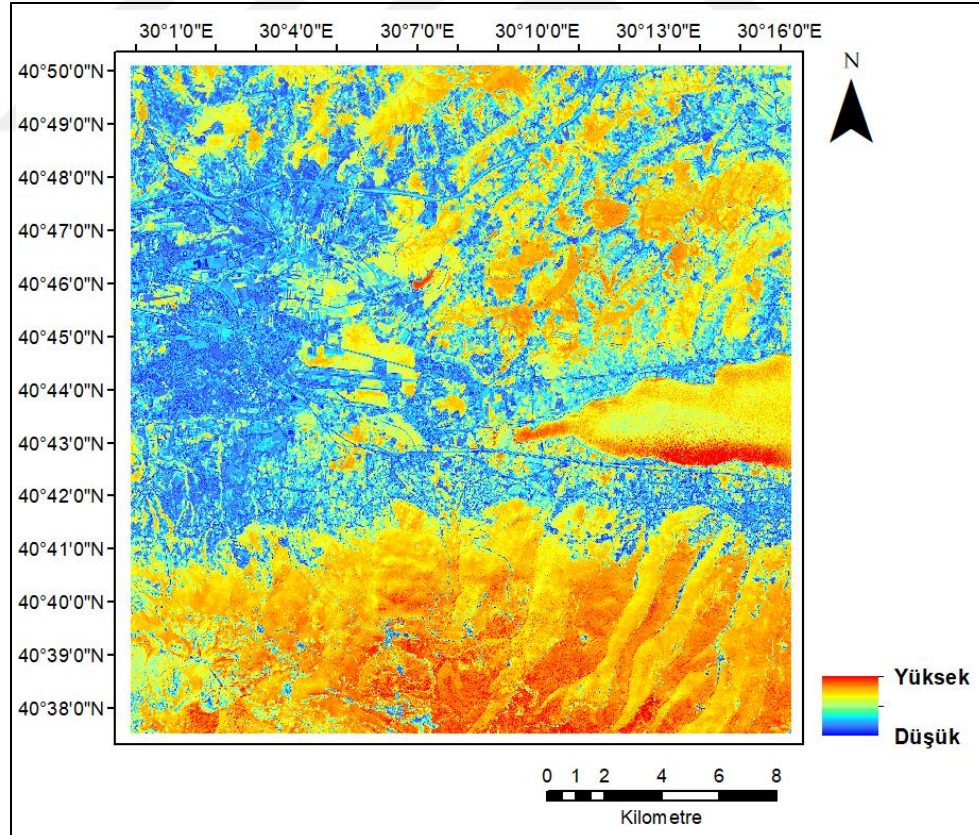
Şekil 5.2: Temel bileşen analizi ile hesaplanan varyans değerleri.

İşlem sonucunda görüldüğü üzere en yüksek varyans değeri ilk temel bileşen ile %62,34 olarak hesaplanmıştır. Bu durum ilk bileşenin diğerlerine göre görüntüde bulunan pikseller açısından daha çok bilgi içerdiğini göstermektedir. İkinci temel bileşenin varyansı %28,80 ve üçüncüsünün %7,28’dir. İlk üç temel bileşenin toplamı

ise %98,42'dir. Üçüncü temel bileşenden sonra varyans değerleri çok düşük değerlere sahip olduğu için sadece ilk üç temel bileşen yardımcı veri olarak değerlendirilmiştir.

5.1.3. Ton-Doygunluk-Yoğunluk Renk Uzay Dönüşümü

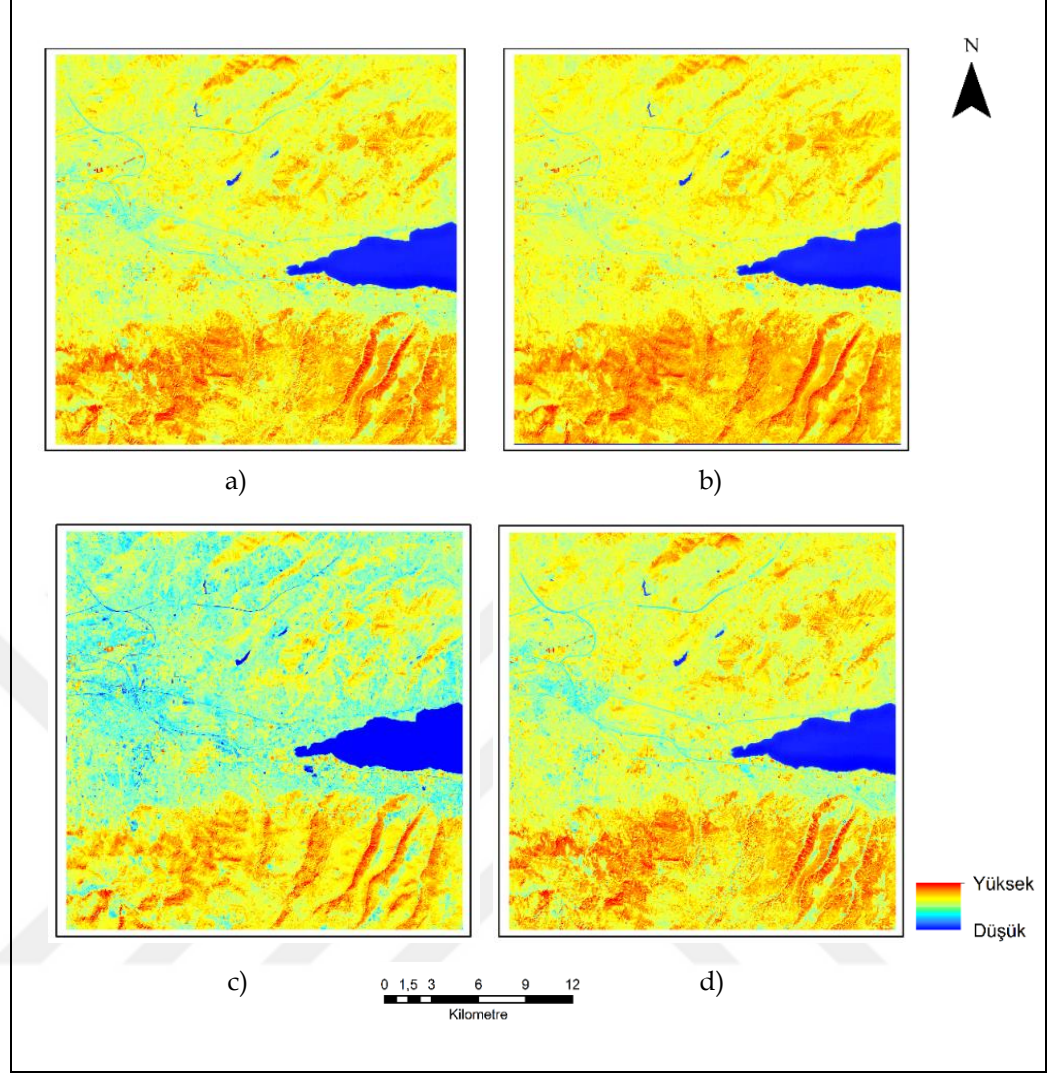
Tez kapsamında değerlendirilen bir diğer yardımcı veri olan ton-doygunluk-yoğunluk (HSI) renk uzay dönüşümüdür. Dönüşümün uygulanması için söz konusu uydu görüntüsünden iki farklı kombinasyon oluşturulmuştur. İlk kombinasyon görüntünün klasik bantları olan kırmızı, yeşil ve mavi bantlardan oluşurken ikinci kombinasyon yakın kızılötesi (NIR), kırmızı ve yeşil bantlardan oluşmuştur. Şekil 5.3'te NIR-R-G bant kombinasyonu için hesaplanmış doygunluk bileşenine ait örnek bir tematik harita verilmiştir. Haritada en yüksek doygunluk değerleri orman ve su piksellerine ait iken en düşük doygunluk değerleri yapılaşmanın olduğu piksellere ait olduğu görülmektedir.



Şekil 5.3: NIR-R-G bant kombinasyonu ile hesaplanan doygunluk haritası.

5.1.4. Morfolojik Filtreler

Yardımcı veri seti oluşturmak için yararlanılan bir diğer görüntü dönüşümü mevcut uydu görüntüsüne morfolojik filtrelerin uygulanması ile elde edilmiştir. Morfolojik işlemler görüntülerde ilgi duyulan nesne veya bölge özelliklerinin belirgin hale getirilmesi veya görüntüde mevcut olan gürültünün ortadan kaldırılması amacıyla uygulanan filtrelerdir. Morfolojik filtre işleminde yapısal eleman (structuring element) olarak adlandırılan matrisler kullanılmaktadır. Bu matrisler görüntü üzerinde dolaştırılarak elde edilmek istenen morfolojik özelliğe göre görüntüdeki pikseller ile karşılaştırılır ve sonuç görüntüsü elde edilir. Morfolojik operatörler genişleme, aşınma, açılma ve kapanma işlemleri olarak adlandırılmaktadırlar. Söz konusu morfolojik operatörler tez çalışması kapsamında kullanılan (mavi, yeşil, kırmızı, kırmızı kenar-1, kırmızı kenar-2, kırmızı kenar-3, NIR, daraltılmış NIR, SWIR-1 ve SWIR-2) spektral bantlara uygulanarak toplamda 40 tane yardımcı veri türetilmiştir. Şekil 5.4'te uydu görüntüsünün NIR bandına, morfolojik operatörler olan genişleme, aşınma, açılma ve kapanma işlemlerinin uygulanmasıyla hesaplanan haritalar gösterilmiştir. Söz konusu şekiller incelendiğinde açılma filtre tekniğinin yol ve şehir alanlarına ait değerlerinin diğer tekniklere göre daha düşük değerler ile hesaplandığı görülmektedir. Açılma filtresi önce genişleme ve ardından aşınma filtresinin görüntülere uygulanmasıyla elde edildiği için böyle bir duruma sebebiyet verdiği söylenilebilir. Ayrıca dört filtre tekniğinde yol, şehir ve su alanlarının değerlerinin düşük, orman gibi ağaç ve bitki yoğunluğunun olduğu alanlar yüksek seviyede değerlere sahip oldukları görülmektedir.

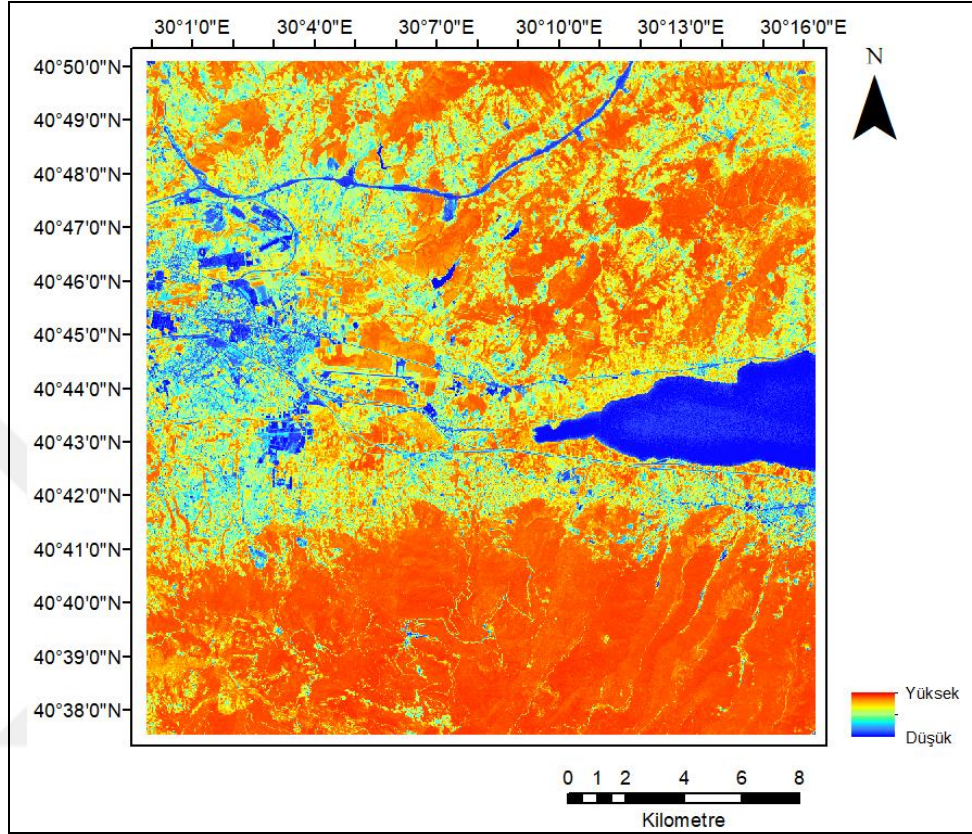


Şekil 5.4: a) Genişleme morfolojik filtresi, b) Aşınma morfolojik filtresi, c) Açılma morfolojik filtresi, d) Kapanma morfolojik filtresi.

5.1.5. Spektral İndisler

Tez kapsamında yararlanılan diğer yardımcı veri spektral indislerdir. Spektral bantların matematiksel eşitliklere bağlı olarak birbirleri ile oranlanması sonucu üretilirler. Spektral indisler, görüntüde farklılık gösteren topoğrafyadan kaynaklanan parlaklık değişimini ortadan kaldırarak verinin renk içeriğinin vurgulanmasını sağlarlar. En bilinenlerden biri olan normalize edilmiş fark bitki indisi (NDVI) haritası Şekil 5.5'te verilmiştir. NDVI indisine ait hesaplanan harita incelenirse canlı bitki örtüsünün diğer nesnelere göre daha yüksek değerler aldığı görülmektedir. Diğer yandan bina, yol, toprak ve su gibi geçirimsiz yüzeylerin daha düşük değerler ile temsil edildiği görülmektedir. Canlı bitki örtüsünün NDVI haritasında daha yüksek değerler

almasının esas nedeni kırmızı dalga boyunda düşük yansımaya değerine sahip olması ve yakın kızılötesi aralığında ise yüksek yansımaya değerine sahip olmasıdır. Hesaplanan indis değerleri çok değişken olduğu için 0 ve 1 arasına normalize edilmiştir.



Şekil 5.5: Çalışma alanına ait NDVI haritası.

Geliştirilmiş ve görüntülere uygulanan birçok indis mevcuttur. Tez kapsamında çeşitli vejetasyon indisleri, su indisleri, toprak indisleri ve yerleşim yeri indisleri olmak üzere toplam 23 tane spektral indis yardımcı veri olarak kullanılmıştır. Kullanılan indisleri şu şekildedir;

Zenginleştirilmiş Bitki Örtüsü İndisi (EVI):

Yoğun bitki örtüsüne sahip alanlarda hassasiyet göstererek bazı atmosferik koşulları ve gölgelik arka plan gürültüsünü düzeltmek için kullanılır [Huete et al., 2002].

$$EVI = 2.5 \times \frac{(NIR - Kırmızı)}{(NIR + 6 \times Kırmızı - 7.5 \times Mavi + 1)} \quad (5.1)$$

Yeşil Atmosferik Dirençli Bitki Örtüsü İndisi (GARI):

GARI çeşitli klorofil yoğunluklarına karşı hassastır ve atmosferik etkilere karşı NDVI'dan daha az duyarlıdır. Gitelson et al. [1996] tarafından atmosferdeki aerosol koşullarına bağlı bir ağırlılandırma fonksiyonu olan gama değerinin (γ) 1,7 olarak işleme alınması önerilmiştir.

$$GARI = \frac{NIR - (Yeşil - \gamma(Mavi - Kırmızı))}{NIR + (Yeşil - \gamma(Mavi - Kırmızı))} \quad (5.2)$$

Yeşil Klorofil İndisi (CIg):

Bitkilerdeki klorofil içeriğine duyarlı olan bu indis, klorofil düzeyinden bitkilerin sağlıklı veya hastalıklı olma durumlarını analiz etmek için kullanılır [Gitelson et al., 2003].

$$GCL = \frac{NIR}{Yeşil} - 1 \quad (5.3)$$

Küresel Çevre İzleme İndisi (GEMI):

Bozucu atmosferik etkilerin azaltılması için üretilmiştir ve doğrusal olmayan bir yaklaşımı esas alır [Pinty & Verstraete, 1992]. Eşitlik 5.4'te verildiği üzere önce bir α değeri bulunur ve sonra GEMI indisi hesaplanır.

$$\alpha = \frac{2 \times (NIR^2 - Kırmızı^2) + 1.5 \times NIR + 0.5 \times Kırmızı}{Kırmızı + NIR + 0.5} \quad (5.4)$$

$$GEMI = \alpha \times (1 - (0.25 \times \alpha)) - \frac{Kırmızı - 0.125}{1 - Kırmızı} \quad (5.5)$$

Yeşil Yaprak İndisi (GLI):

Bitkileri, toprak veya diğer nesnelerden ayırt edebilen bir spektral indistir ve -1 ile +1 arasında değer almaktadır. Pozitif değerler yeşil canlı bitkiyi negatif değerler ise canlı olmayan nesneleri temsil etmektedir [Louhaichi et al., 2001]. GLI indisi Eşitlik 5.6'da verilen formülle hesaplanmaktadır.

$$GLI = \frac{(2 \times Yeşil) - Kırmızı - Mavi}{(2 \times Yeşil) + Kırmızı + Mavi} \quad (5.6)$$

Yeşil Normalize Edilmiş Fark Bitki Örtüsü İndisi (GNDVI):

Bitkilerdeki fotosentetik aktiviteyi kontrol etmek amacıyla geliştirilmiştir [Gitelson et al., 1998]. NDVI'dan farkı kırmızı bant yerine yeşil bandın kullanılmasıdır ve belirtilen Eşitlik 5.7 ile hesaplanır.

$$GNDVI = \frac{NIR - Yeşil}{NIR + Yeşil} \quad (5.7)$$

Kızılötesi Oran Bitki Örtüsü İndisi (IPVI):

IPVI fonksiyonel olarak NDVI'a benzemektedir fakat hesaplanması daha hızlıdır, ayrıca NDVI'dan farklı olarak 0 ve 1 aralığında değerler almaktadır. [Crippen, 1990]. Eşitlikte 5.8'de verildiği üzere yakın kızıl ötesi ve kırmızı bant kullanılarak hesaplanır.

$$IPVI = \frac{NIR}{NIR + Kırmızı} \quad (5.8)$$

Nem Stres İndisi (MSI):

Bitkinin yaprağındaki su miktarına göre farklı yansıtma özelliği gösterir ve MSI ile bu özellikten faydalanılarak bitkinin su miktarı tespit edilmektedir [Hunt & Rock, 1989]. Eşitlik 5.9’da verilen formülde hangi dalga boylarını kullanılarak hesaplandığı belirtilmiştir.

$$MSI = \frac{SWIR}{NIR} \quad (5.9)$$

Normalize Edilmiş Fark Bitki Örtüsü İndisi (NDVI):

Bitkinin güçlü şekilde yansıttığı ve bitkinin soğurduğu dalga boyu bölgeleri (NIR, Kırmızı) kullanılarak hesaplanır. Sağlıklı bitki örtüsü +1’e yakın değerler alırken sağlıklı olmayan veya cansız nesneler -1’e yakın değerler ile temsil edilir [Goward et al., 1991].

$$NDVI = \frac{NIR - Kırmızı}{NIR + Kırmızı} \quad (5.10)$$

Optimize Edilmiş Toprak Ayarlı Bitki Örtüsü İndisi (OSAVI):

Bu indis SAVI’nin optimize edilmiş halidir ve Eşitlik 5.11’de verildiği üzere kanopi arka plan için sabit değer olarak 0.16 kullanıldığında tarımsal alanlarda başarılı sonuçlar elde edilmektedir [Rondeaux et al., 1996].

$$OSAVI = \frac{NIR - Kırmızı}{NIR + Kırmızı + 0.16} \quad (5.11)$$

Yeniden Normalize Edilmiş Bitki Örtüsü İndisi (RDVI):

Sağlıklı bitki örtüsünün tespit edilmesi için NDVI’da olduğu gibi yakın kızılötesi ve kırmızı dalga boyları kullanılarak üretilir [Roujean et al., 1995]. Eşitlik 5.12’de verildiği üzere NDVI’dan farklı olarak payda kısmı karekök alınarak hesaplanmaktadır.

$$RDVI = \frac{NIR - Kırmızı}{\sqrt{NIR + Kırmızı}} \quad (5.12)$$

Kırmızı Kenar Normalize Edilmiş Fark Bitki İndisi (RENDVI):

RENDVI esas yutulma ve yansıtma tepe noktaları yerine kırmızı kenar (Red Edge) bantlarının kullanılmasıyla hesaplanır ve NDVI’nın bir biçimidir [Gitelson and Merzlyak, 1994].

$$RENDVI = \frac{NIR - Kırmızı Kenar}{NIR + Kırmızı Kenar} \quad (5.13)$$

Toprak Ayarlı Toplam Bitki Örtüsü İndisi (SATVI):

Yeşil ve kuru bitki örtüsünün belirlenmesi amacıyla (NDSVI, SAVI) iki indisin bir bileşimi olarak üretilmiştir [Marsett et al., 2006]. Bu indis Eşitlik 5.14 ile verilen formülle hesaplanmaktadır.

$$SATVI = \left(\frac{SWIR1 - Kırmızı}{SWIR1 + Kırmızı + 0.5} \right) \times 1.5 - \frac{SWIR2}{2} \quad (5.14)$$

Toprak Ayarlı Bitki Örtüsü İndisi (SAVI):

Bitki örtüsünün miktarının az olması halinde topraktan kaynaklanan parlaklığın etkisini gidermek için NDVI'ın değiştirilmiş bir hali olarak üretilmiştir [Huete, 1988]. Eşitlik 5.15'te verilen formülle hesaplanmaktadır.

$$SAVI = 1.5 \times \frac{NIR - Kırmızı}{NIR + Kırmızı + 0.5} \quad (5.15)$$

Yapısal Duyarsız Pigment İndisi (SIPI):

Bitki klorofil oranının tahmin edilmesi için kullanılan SIPI, 0 ile 2 arasında değer alır ve sağlıklı bitki değeri 0.8 ile 1.8 arasındaki değerlerdir [Penuelas et al., 1995]. Eşitlik 5.16'te nasıl hesaplandığı gösterilmiştir.

$$SIPI = \frac{NIR - Mavi}{NIR - Kırmızı} \quad (5.16)$$

Dönüştürülmüş Fark Bitki Örtüsü İndisi (TDVI):

Kentsel alanlarda bitki örtüsünün izlenmesi için üretilmiştir ve NDVI ile SAVI gibi doygunluk göstermemektedir [Bannari et al., 2002]. 5.17 ile verilen formülle hesaplanmaktadır.

$$TDVI = 1.5 \times \left[\frac{NIR - Kırmızı}{\sqrt{NIR^2 + Kırmızı} + 0.5} \right] \quad (5.17)$$

Görünür Atmosferik Direnç İndisi (VARI):

Atmosferik etkilere karşı düşük hassasiyete sahip bir görüntüdeki bitki örtüsünün fraksiyonunu tahmin etmek için kullanılır ve ARVI indisini baz alarak üretilmiştir [Gitelson et al., 2002]. Eşitlik 5.18 olarak verilen denklemde nasıl hesaplandığı verilmiştir.

$$VARI = \frac{Yeşil - Kırmızı}{Yeşil + Kırmızı - Mavi} \quad (5.18)$$

Normalize Edilmiş Fark Su İndisi (NDWI):

Gao [1996] tarafından bitkilerde ve toprakta bulunan nem miktarının belirlenmesi amacıyla geliştirilmiştir. Ayrıca NDWI su kütlelerin belirlenmesi amacıyla görüntülere uygulanmaktadır. Eşitlik 5.19’da verilen denklem ile hesaplanmaktadır.

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR} \quad (5.19)$$

Su Oranı İndisi (WRI):

Su ve buz karakteristiklerine sahip alanlar ile karasal alanların birbirlerinden düzgün bir şekilde ayırt edilebilmesi için geliştirilmiştir [Shen & Li, 2010]. Eşitlik 5.20’de belirtildiği üzere orta yeşil, kırmızı, yakın ve orta kızılötesi bantlar ile hesaplanmaktadır.

$$WRI = \frac{Yeşil + Kırmızı}{NIR + MIR} \quad (5.20)$$

Yerleşim Alanı Çıkarım İndisi (BAEI):

BAEI, görüntüdeki yerleşim yerlerinin vurgulanması amacıyla geliştirilmiştir [Bouzekri et al., 2015]. Denklemden görüldüğü üzere 0.3 olarak belirlenmiş bir L aritmetik sabiti kullanılmaktadır.

$$BAEI = \frac{Kırmızı + L}{Yeşil + SWIR} \quad (5.21)$$

Yerleşim Alanı için Bant Oranı (BRBA):

Kırmızı ve kızılötesi bantların kullanımıyla çıplak toprak ve yerleşim yerlerinin birbirlerinden ayırt edilmesi için geliştirilmiştir [Waqar et al., 2012]. Belirtilen Eşitlik 5.22 ile hesaplanır.

$$BRBA = \frac{MIR}{Kırmızı} \quad (5.22)$$

Kuru Çıplak-Toprak İndisi (DBSI):

Çıplak alanların ve yerleşim yerlerinin tespit edilmesi için geliştirilmiştir. -2 ile 2 arasında değerler almaktadır ve 0.26'dan büyük değerler çıplak toprak olarak değerlendirilmektedir [Rasul et al., 2018]. DBSI indisi Eşitlik 5.23 ile hesaplanmaktadır.

$$DBSI = \left(\frac{SWIR1 - Yeşil}{SWIR1 + Yeşil} \right) - NDVI \quad (5.23)$$

Normalize Edilmiş Fark Toprak Sürme İndisi (NDTI):

Toprak alanları ve yerleşim yerlerinin birbirlerinden ayrılması amacıyla önerilmiştir [Van Deventer et al., 1997]. NDTI indisi Eşitlik 5.24 ile hesaplanmaktadır.

$$NDTI = \frac{SWIR1 - SWIR2}{SWIR1 + SWIR2} \quad (5.24)$$

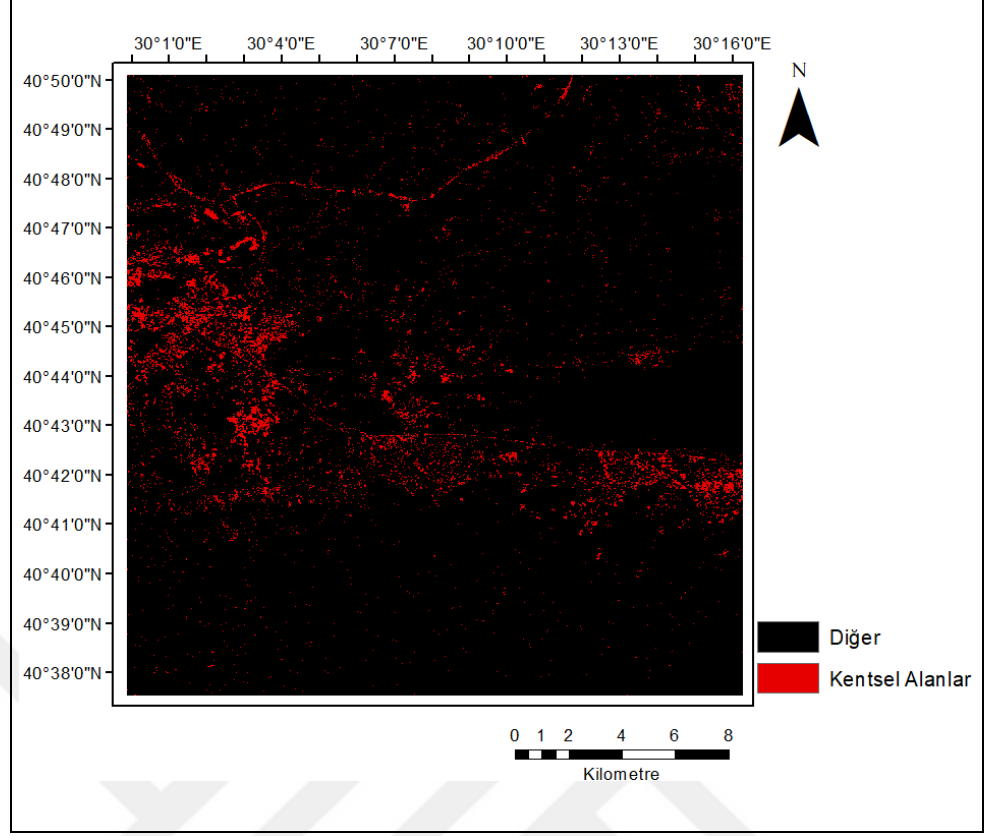
5.2. Diğer Yardımcı Veriler

Uydu görüntüsünün bantları kullanılarak türetilen yardımcı verilerin (ancillary data) yanında başka bir dış kaynaktan edinilen çeşitli yardımcı verilerde (auxiliary data) tez çalışmasında kullanılmıştır. Sınıflandırma işleminin ne ölçüde

etkileneceğinin analizinin sağlıklı bir şekilde yapılması için dış kaynaklardan farklı şekillerde yardımcı veri türleri elde edilmiştir. Bu veriler çalışma alanına ait küresel kentsel ayak izi haritası, nüfus haritası, sayısal yükseklik modeli ile üretilen eğim, bakı ve yükseklik haritaları, yeryüzüne ait toprak türü haritası, yağış miktarı ve toprak sıcaklığını temsil eden haritalardır. Bu yardımcı verilerin özellik seçim işlemlerinde değerlendirilmesiyle, sınıflandırma sürecine bir katkıları varsa bunun ne ölçüde olduğunun araştırması yapılmıştır.

5.2.1. Küresel Kentsel Ayak İzi

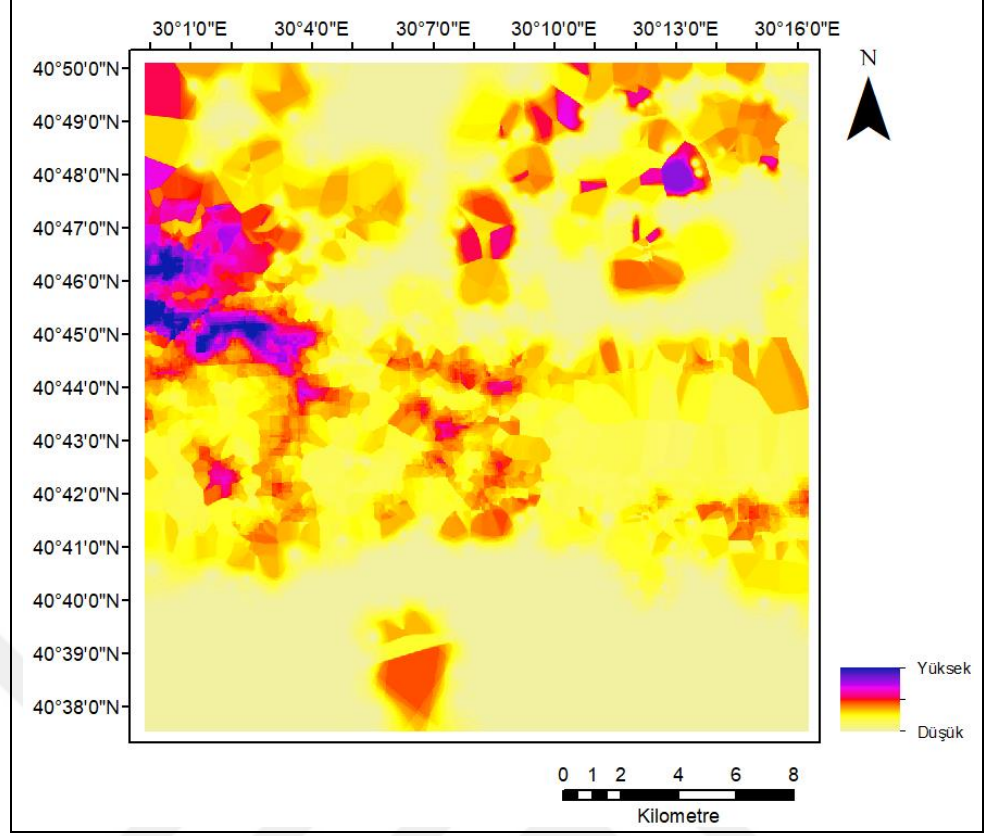
Dış kaynaktan türetilen ilk yardımcı veri küresel kentsel ayak izi (GUF) haritasıdır. İnsan yerleşmelerinin gelişiminin dinamiklerini ve etkilerini çıkarmak için kullanılan haritalardır. Bu nedenle GUF haritaları, insan yerleşimlerinin konumu ve dağılımının güncel bir şekilde mekânsal olarak temsil eden ürünlerdir denilebilir [Esch et al., 2017]. Bu tez çalışması kapsamında GUF haritası üretebilmek için 29 Ocak 2021 ve 2 Nisan 2021 tarihlerine ait, çalışma alanını kapsayan Sentinel-1A yapay açıklıklı radar görüntüleri (SAR) interferometrik geniş alan (IW) modunda ESA resmi sitesinden temin edilmiştir. Çalışma alanına ait GUF haritasını üretmek için her iki görüntüye SNAP (Sentinel Application Platform) yazılımında radyometrik kalibrasyon işlemi uygulanmıştır. Bu işlemten sonra veri olmayan değerlerin ortadan kaldırılması ve görüntüde var olan yatay şeritlerin giderilmesi (deburst) işlemi iki görüntü için gerçekleştirilmiştir. Görüntülerde benek etkisinin azaltılması ve piksellerin x ve y eksenlerinin aynı boyuta getirilmesi için multilooking (çok bakışlı) adı verilen SNAP işlemi gerçekleştirilmiştir. Bu işlemten sonra her iki görüntü için SYM verisi kullanılarak geometrik düzeltme işlemi Range-Doppler arazi düzeltmesi ile yapılmıştır. Daha sonra uydu görüntüleri arasındaki uyum (coherence) ve genlik hesaplanmıştır. Sonra iki görüntü istif haline getirilmiştir. Görüntülerin ortalaması alınarak aralarındaki farklılık belirlenmiş ve GUF haritası üretilmiştir. Üretilen harita ArcGIS yazılımında analiz edilmiş ve diğer yardımcı veriler ile etkili bir şekilde değerlendirme yapılabilmesi için mekânsal çözünürlüğü en yakın komşuluk yöntemiyle 12,5 metreye yeniden örneklendirilmiştir. Şekil 5.6'da görüldüğü üzere GUF haritasında belirtilen kırmızı pikseller çalışma alanındaki yerleşim alanlarını siyah ile temsil edilenler diğer arazi sınıflarını belirtmektedir.



Şekil 5.6: Çalışma alanına ait küresel kentsel ayak izi haritası.

5.2.2. Nüfus

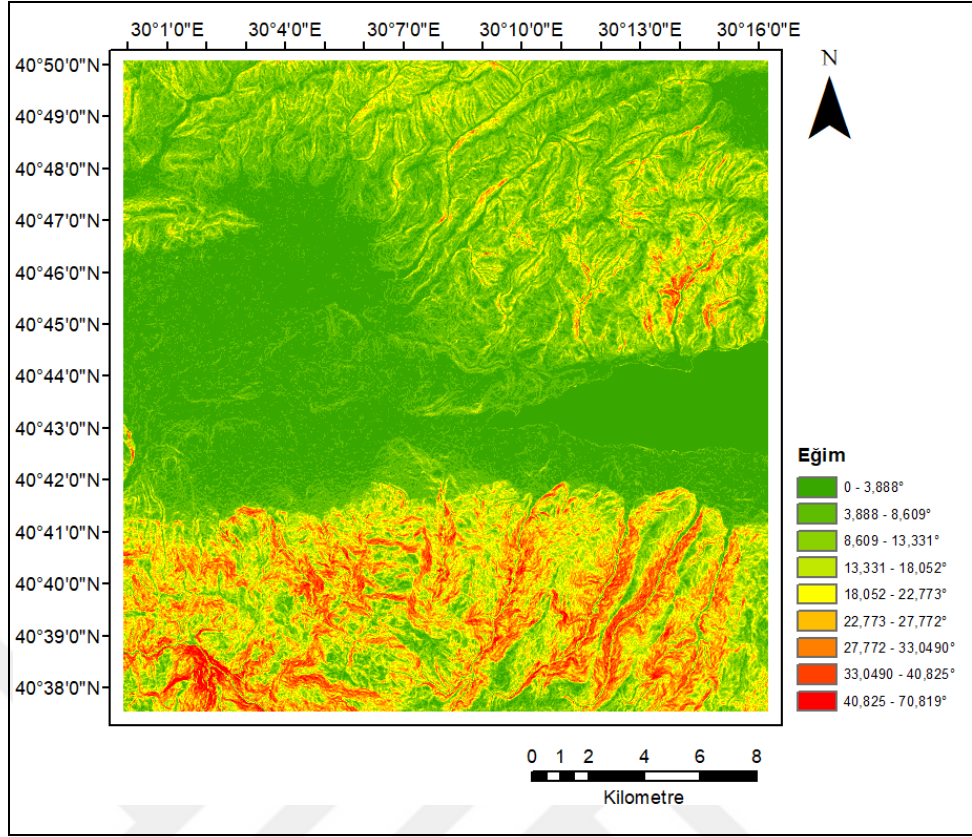
Diğer yardımcı veri her bir pikselde bulunan insan sayısının temsil edildiği nüfus haritasıdır. Diğerlerine göre yüksek nüfus yoğunluğuna sahip sınıflar için tahmin doğruluğunu artırmak ve kentsel yerleşimlere spektral olarak benzer sınıflar arasındaki karışıklığı azaltmak için küresel nüfus veri seti kullanılmaktadır [Hurskainen et al., 2019]. Bu veri, nüfusun mekânsal olarak dağıtılmış, 100 metre çözünürlüğe sahip güncel ve küresel raster veri sağlayan bir açık kaynaktan temin edilmiştir [Lloyd et al., 2017], [Web 4, 2021]. Kullanılan nüfus haritası 2020 yılına aittir ve veride nüfusun olmadığı alanlarda boş değerler mevcuttur. Diğer yardımcı veriler ile analiz edilebilmesi için ters mesafe ağırlıklı mesafe yöntemi (IDW) kullanılarak boş alanlar için değerler hesaplanmıştır. Verinin mekânsal çözünürlüğü pikseller ile temsil edilen nüfus sayıları dikkate alınarak en yakın komşuluk yöntemi ile 12.5 metreye yeniden örneklendirilmiş. Bu işlem ile elde edilen nüfus yoğunluğuna ait tematik harita Şekil 5.7 ile gösterilmiştir.



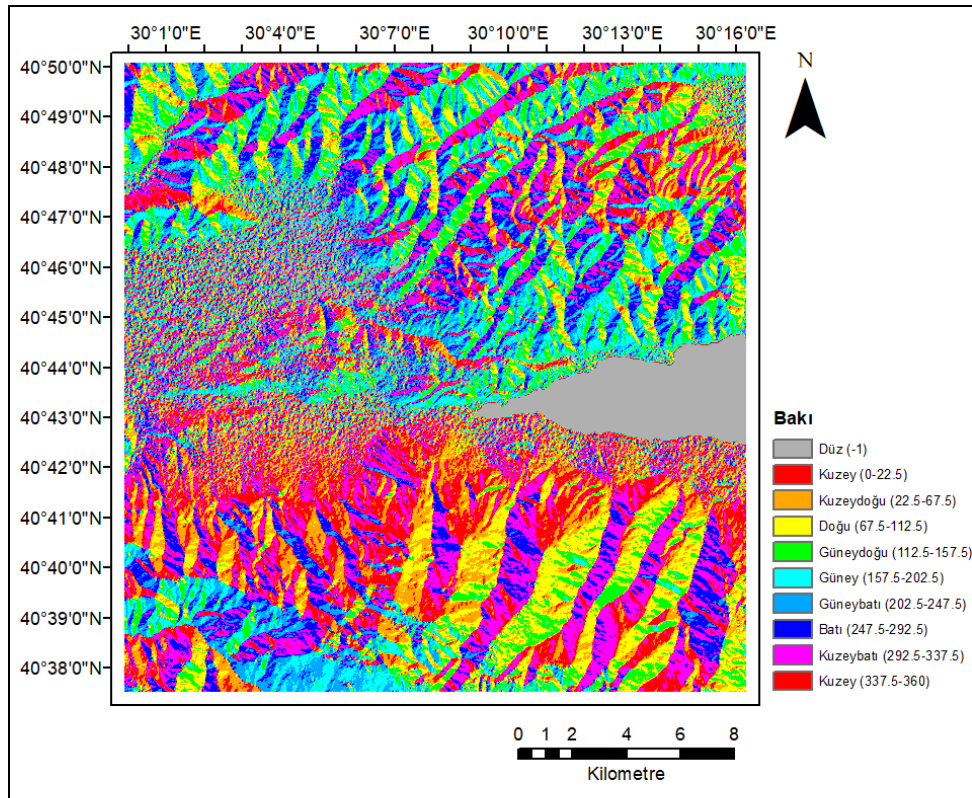
Şekil 5.7: Çalışma alanına ait nüfus haritası.

5.2.3. Eğim, Bakı ve Yükseklik

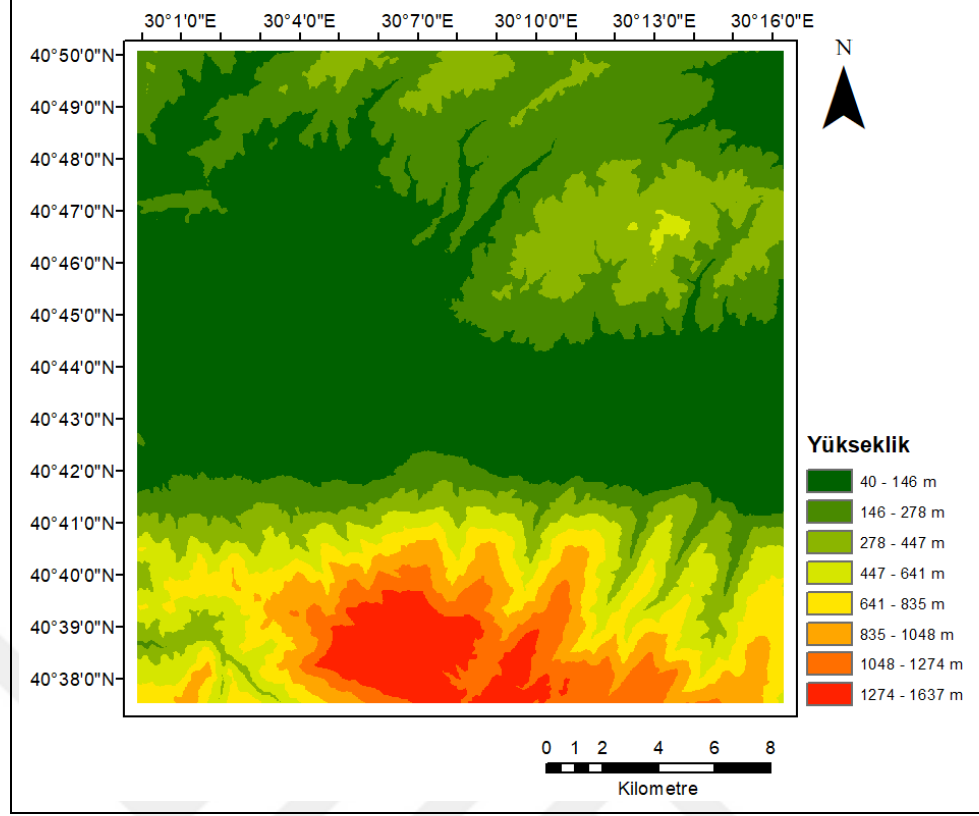
Özellik seçimi analizinin derinleştirilmesi için elde edilen bir diğer yardımcı veri sayısal yükseklik modelinden (SYM) elde edilen yükseklik, eğim ve bakıdır. Çünkü yüksek topografik değişikliklerin olduğu alanlarda aynı sınıfa ait piksellerin farklı spektral değerlere ya da farklı sınıflara ait piksellerin benzer spektral değerlere sahip olması söz konusu olabilmektedir [Fahsi et al., 2000]. SYM verisi ALOS-PALSAR uydusundan temin edilerek yükseklik, eğim ve bakı haritaları oluşturulmuştur. Veri 12.5 metre mekânsal çözünürlükte ve ücretsiz erişime açıktır. Şekil 5.8’de gösterildiği üzere çalışma alanında 0° ile 70° arasında değişen eğim değerleri bulunmaktadır. Şekil 5.9’da çalışma alanına ait bakı haritası ve Şekil 5.10’da çalışma alanına ait 40 ile 1637 m arasında değişen topoğrafik yükseklik haritası verilmiştir.



Şekil 5.8: Çalışma alanına ait eğim haritası.



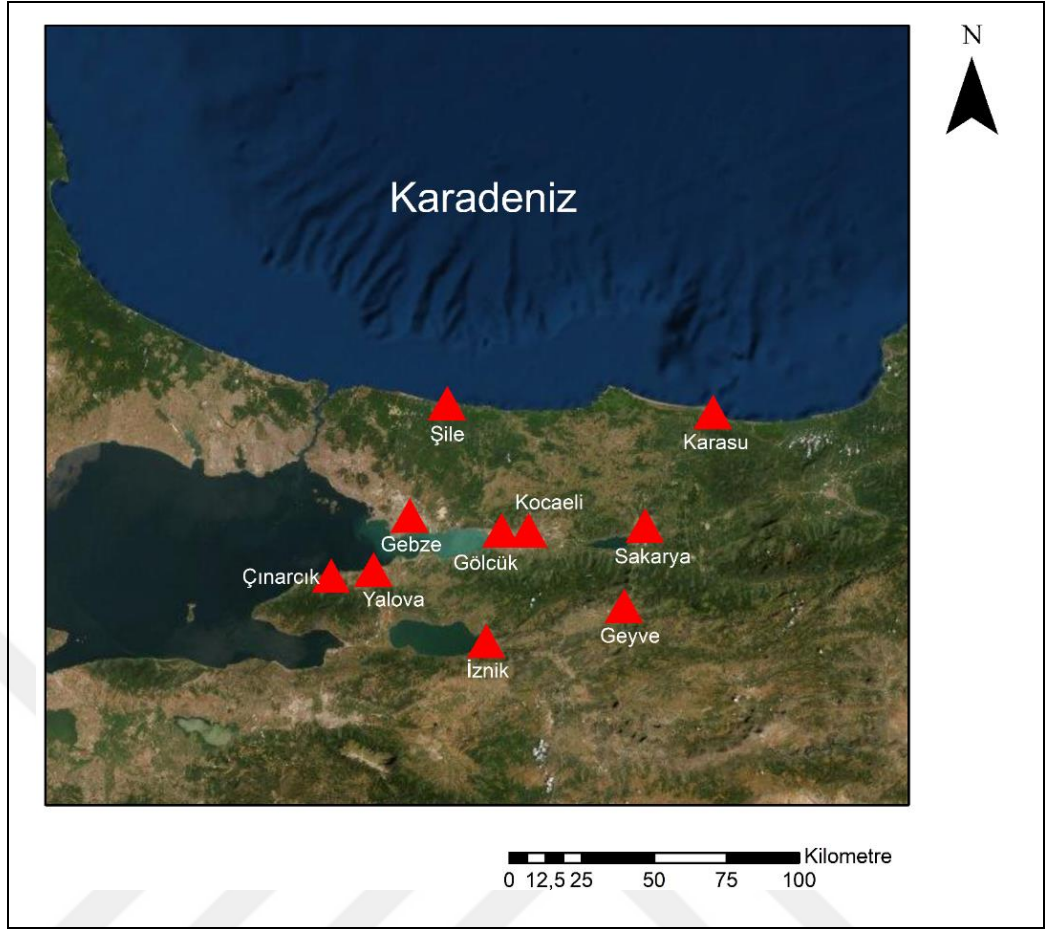
Şekil 5.9: Çalışma alanına ait bakı haritası.



Şekil 5.10: Çalışma alanına ait yükseklik haritası.

5.2.4. Toprak Sıcaklığı

Arazi örtüsü tipi, genellikle zemin yüzeyleri ile yerel sıcaklık arasındaki etkileşimi anlamak için kullanılan önemli bir özelliktir [İbrahim et al., 2016]. Toprak sıcaklığı özellikle üzerinde yetişen bitki örtü türünü ise doğrudan etkilemektedir. Bu nedenle bu çalışmada bir başka yardımcı veri olan toprak sıcaklığı bir özellik olarak kullanılmıştır. Toprak sıcaklık verileri T.C. Tarım ve Orman Bakanlığı Meteoroloji Genel Müdürlüğü'nden temin edilmiştir. Toprak sıcaklık verisi 5 cm'lik toprak kalınlığına aittir ve Şekil 5.11'de verildiği üzere 10 farklı istasyonun, 2019 yılına ait aylık toprak sıcaklıklarının ortalaması hesaplanmıştır. Hesaplanan değerler ile IDW yöntemi kullanılarak toprak sıcaklık haritası üretilmiştir. Böylece toprak sıcaklığının uydu görüntüsü sınıflandırmada etkisinin incelenmesi amaçlanmıştır.



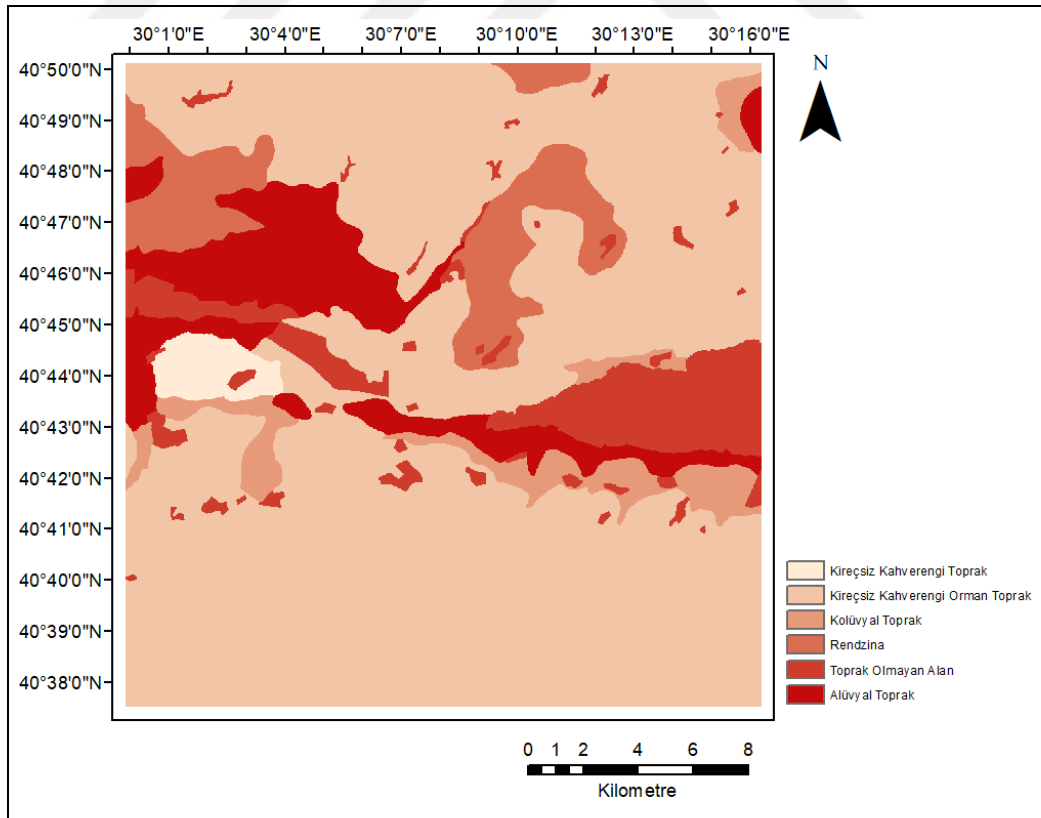
Şekil 5.11: Verilerin temin edildiği istasyonlar.

5.2.5. Yağış Miktarı

T.C. Tarım ve Orman Bakanlığı Meteoroloji Genel Müdürlüğü'nden temin edilen diğer yardımcı veri ise yağış miktarıdır. Yağış miktarı özellikle bitki gelişiminde önemli bir rol oynadığı için yardımcı veri olarak tez kapsamında değerlendirilmiştir. 2019 yılına ait aylık yağış miktarının (kg/m^2) ortalaması hesaplanmıştır. Şekil 5.11'de verilen istasyonlardan alınan veriler enterpolasyon işlemi için kullanılmıştır. Hesaplanan değerlere IDW yöntemi uygulanarak yağış miktarını temsil eden harita üretilmiş ve yardımcı veri olarak çalışmada değerlendirilmiştir. Bu yardımcı verinin değerlendirilmesi ile yağış miktarının arazi örtüsü/arazi kullanımı haritası üretme aşamasında önemli bir özellik olup olmadığının analizi yapılması amaçlanmıştır.

5.2.6. Toprak Türü

Çalışma alanına ait, yeryüzü toprak türü özellikleri sınıflandırma sürecinde değerlendirmeye alınan bir diğer özelliktir. Toprak çeşitliliği bitki örtüsünün dağılımını haritalamak için kullanılan bir yardımcı veridir [Zeferino et al., 2020]. Çalışma alanına ait toprak türlerinin yardımcı veri olarak uydu görüntüsü sınıflandırılmasında etkinliğinin analiz edilebilmesi için Tarım ve Orman Bakanlığı'ndan toprak türlerine ait veri temin edilmiştir. Verilerle birlikte analiz edilebilmesi için mekânsal çözünürlüğü diğer işlemlerde olduğu gibi en yakın komşuluk yöntemi kullanılarak 12,5 metreye yeniden örneklendirilmiştir. Çalışma alanında bulunan toprak türleri kireçsiz kahverengi topraklar, kireçsiz kahverengi orman topraklar, kolüvyal topraklar, rendzinalar, alüvyal topraklar ve kahverengi orman topraklarıdır. Şekil 5.12'de toprak türlerinin ve arazide nasıl bir dağılım gösterdikleri gösterilmiştir. Toprak türleri incelendiğinde çalışma alanında en fazla bulunan toprak türü kireçsiz kahverengi orman toprağıdır.



Şekil 5.12: Çalışma alanına ait toprak türü haritası.

6. UYGULAMA

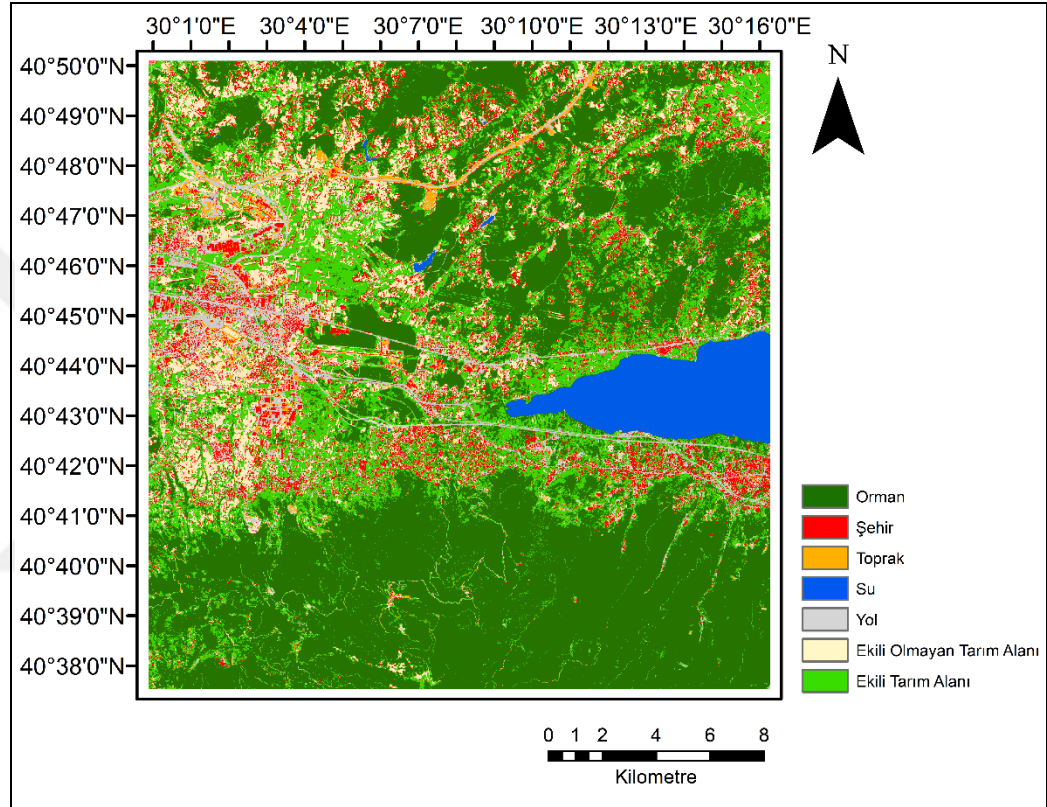
Sentinel-2A uydu görüntüsü ve yardımcı verilerin birleştirilmesiyle elde edilen 170 özellikten oluşan yüksek boyutlu veri setinden hangi özelliklerin arazi sınıf ayrımını en iyi şekilde yapabildiğinin belirlenmesi için detaylı bir analiz gerçekleştirilmiştir. İlk olarak mevcut Sentinel-2A uydu görüntüsünün spektral bantları ile sınıflandırma işlemleri yapılmıştır. Bu işlemin amacı boyut indirgeme ile elde edilen sonuçlar ile karşılaştırılarak yardımcı veri setinin görüntü sınıflandırma işleminde etkinliğinin belirlenmesidir. Sınıflandırma sürecinde faydasız, korelasyonlu ve tanımlamaya katkı sağlamayan bantların elenerek boyut indirgeme işlemlerinin yapılabilmesi için özellik seçim yöntemleri kullanılmıştır. Bu amaç doğrultusunda üç filtreleme özellik seçim yöntemi (Ki-kare, ReliefF ve Bilgi Kazancı), üç sarmal özellik seçim yaklaşımı (GA, AIS, ÖÖS) ve son olarak üç gömülü özellik seçim yöntemi (LightGBM, HSIC Lasso ve RO) oluşturulan veri setlerine uygulanmıştır. Sınıflandırma işlemine en çok katkı sağlayan bantların belirlenmesi amacıyla söz konusu dokuz algoritma tarafından tespit edilen alt kümeler dikkate alınmıştır. Her bir özellik seçim yöntemi ile elde edilen özelliklerle 10, 20, 30, 40, 50, 75, 100, 135 ve 170 banttan oluşan dokuz farklı veri seti oluşturulmuştur. Belirlenen bu veri setlerine, rastgele orman algoritması uygulanarak genel doğrulukları hesaplanmış ve elde edilen en yüksek genel doğruluğa sahip veri setleri rastgele orman (RO), LightGBM, XGBoost (XGB) ve oylama (voting) algoritmaları ile değerlendirilip sınıflandırma performansları karşılaştırılmıştır. Özellik seçim yöntemleriyle belirlenen tüm veri setleri için, çalışma kapsamında RO algoritması kullanılmış ve ormandaki ağaç sayısı 100 olarak belirlenmiştir. Ayrıca RO yönteminin diğer parametreleri scikit-learn kütüphanesinde varsayılan olarak seçilmiştir. XGB algoritması için öğrencinin türünün ayarlandığı parametre, ağaç topluluğunu oluşturan 'gbtree' olarak seçilmiş ve ağaç sayısı ise 100 olarak belirlenip tüm aynı şekilde işlemlerde kullanılmıştır. Çalışmada yardımcı verilerin değerlendirilmesiyle veri boyutundaki artışın sınıflandırma performansına olan etkisinin ortaya konulması ve detaylı bir analizin yapılması hedeflenmiştir. Özellik seçim işlemlerinin sonucunda 72 farklı eğitim veri seti (dokuz özellik seçimi yöntemine göre belirlenen sekiz veri seti) üretilmiş ve detaylı analizleri yapılmıştır.

6.1. Uydu Görüntüsünün Sınıflandırılması

Çalışmaya konu olan uzaktan algılanmış mevcut Sentinel-2A uydu görüntüsü için, ek veri kullanılmadan 10 spektral bandının değerlendirilmesiyle sınıflandırma işlemleri gerçekleştirilmiştir. Sınıflandırma işlemlerinde 10 m ve 20 m mekânsal çözünürlüğe sahip spektral bantlar 12,5 m mekânsal çözünürlüğe yeniden örneklendirilmiştir. Yeniden örnekleme işleminin yapılmasının amacı 12,5 m mekânsal çözünürlüğe sahip olan yardımcı verilerin (eğim, bakı ve yükseklik) eklenmesiyle oluşturulan yüksek boyutlu görüntü ile sınıflandırma sonuçlarının karşılaştırılabilmesidir. Yapılan sınıflandırma işlemlerinde RO, DVM ve XGB yöntemleri ve bu üç yöntemin ele alındığı oylama tekniği değerlendirilmiştir. Tablo 6.1’de sınıflandırma sonuçlarına ait genel doğruluklar ve kullanılan makine öğrenme yöntemi verilmiştir. Sonuçlar incelendiğinde en yüksek doğruluğun söz konusu üç algoritmanın birlikte kullanıldığı oylama yöntemi ile hesaplandığı görülmektedir. Bu sınıflandırma işleminin yapılmasındaki asıl amaç; yardımcı verilerin ve boyut indirgeme yöntemlerinin değerlendirilmesiyle gerçekleştirilen sınıflandırma işlemlerinin, uydu görüntüsünün mevcut bantlarının kullanılmasıyla yapılan sınıflandırma işlemlerinden üstün bir performans gösterip göstermediği belirlemektir. Bu nedenle mevcut spektral bantlar ile yapılan sınıflandırma işleminin sonuçlarını hesaplamak önem arz etmektedir. Çünkü tez çalışmasında bu işlem, yardımcı veriler ile oluşturulan yüksek boyutlu görüntüye uygulanan ve çeşitli temellere dayanan boyut indirgeme algoritmalarını değerlendirmek için bir dayanak durumundadır. Elde edilen veri setlerinin sınıflandırma sonuçlarına nasıl etki ettiği çalışmada önemli bir çıktıdır. Şekil 6.1’de uydu görüntüsünün spektral bantlarının oylama sınıflandırma sonucunda üretilen tematik haritası gösterilmiştir. Bu sınıflandırma işleminde ve diğer tüm uygulamalarda kullanılan RO ve XGB algoritmaları için ağaç sayısı 100 olarak belirlenmiş ve XGB için öğrenicinin türünün belirlendiği parametre “gbtree” olarak seçilmiştir.

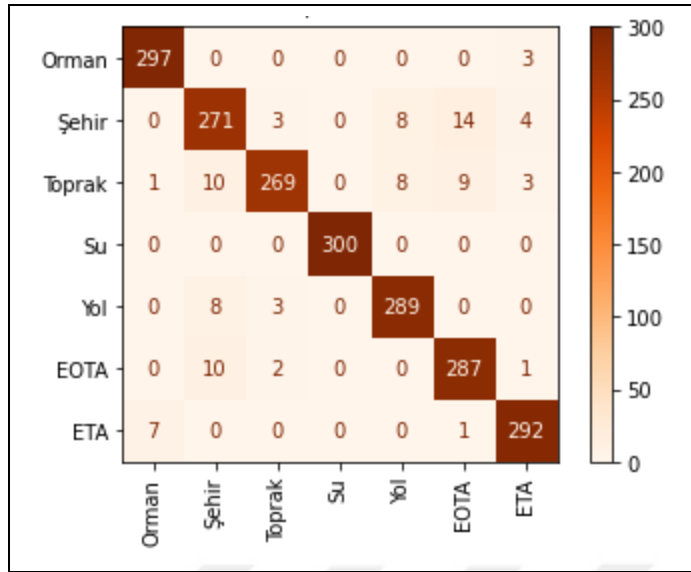
Tablo 6.1: Uydu görüntüsü sınıflandırma genel doğrulukları.

Algoritma	Genel Doğruluk (%)
RO	94,86
XGB	94,33
DVM	93,86
Oylama	95,48



Şekil 6.1: Sentinel-2A uydu görüntüsünün oylama tekniğiyle sınıflandırılması sonucu üretilen tematik harita.

Uydu görüntüsünün rastgele orman yöntemiyle sınıflandırma işlemi sonucunda hesaplanan hata matrisi Şekil 6.2’de gösterilmiştir. Hata matrisinde, satırlarda yer alan değerler referans piksellerini ve sütunlardaki değerler ise tahmin edilen pikselleri temsil etmektedir. Hata matrisi incelendiğinde toprak sınıfının diğer arazi sınıflarına göre daha kötü bir performans ile sınıflandırıldığı görülmektedir. Ayrıca toprak sınıfındaki pikseller, şehir ve ekili olmayan tarım alanı sınıflarıyla daha fazla karıştığı gözlemlenmektedir. Hata matrisinde toprak sınıfı en az orman ve su sınıfı ile karıştığı görülmektedir. Bunun nedeni ise toprağı ve söz konusu diğer sınıfları temsil eden piksellerin spektral yansımalarının birbirinden farklı olmasından kaynaklanmaktadır.



Şekil 6.2: Sentinel-2A uydu görüntüsünün oylama tekniğiyle sınıflandırılması sonucu üretilen hata matrisi.

6.2. Filtre Tabanlı Boyut İndirgeme

Ki-kare, ReliefF ve Bilgi Kazancı olarak üç farklı filtre özellik seçim yöntemi 170 banttıan oluřan çok boyutlu görüntüye uygulanmıřtır. Bunun sonucunda 10, 20, 30, 40, 50, 75, 100 ve 135 banttıan ve tüm veri setinden oluřan alt kümeler elde edilmiřtir. Bu alt kümeler sırasıyla VS-1, VS-2, VS-3, VS-4, VS-5, VS-6, VS-7, VS-8, VS-9 olarak adlandırılmıřtır. Tüm özellik seçim iřlemleri Python 3.8.3 programlama dilinde ve Anaconda'nın Jupyter Notebook ara yüzü kullanılarak gerekleřtirilerek analizleri yapılmıřtır. Ki-kare yöntemi kullanılarak gerekleřtirilen özellik seçim iřlemi 6,5 milisaniye, ReliefF algoritması ile yapılan özellik seçim iřlemi 9,04 saniye ve bilgi kazancı özellik seçim yönteminin uygulanmasıyla özellik seçim iřlemi 3,83 saniye sürmüřtür. Boyut indirgeme iřlemleri iin gerekli olan zaman maliyetlerinden anlařılacađı üzere söz konusu filtre tabanlı özellik seçim yöntemleri ile hızlı bir řekilde boyut indirgeme iřlemi gerekleřtirildiđi sonucuna ulařılmıřtır. Ayrıca en hızlı yöntem Ki-kare olurken diđerlerine göre daha yavař filtre tabanlı özellik seçim yöntemi ReliefF olmuřtur. Buna rađmen üç yöntem arasındaki zaman maliyet farkı saniyeler mertebesinde dir.

6.2.1. Ki-kare Yöntemi ile Boyut İndirgeme

Ki-kare boyut indirgeme yöntemi kullanılarak hangi bantların sınıflandırma işlemine daha çok katkı sağlayacağını tespit etmek amacıyla 170 banttan oluşan tüm veri setinin yanı sıra özellik seçim işlemi sonucunda farklı boyutlarda sekiz veri seti belirlenmiştir. Söz konusu veri setlerinin sınıflandırma işlemiyle hesaplanan genel doğruluklarına göre en yüksek performansın VS-7'ye ait olduğu tespit edilmiştir. 100 banttan oluşan bu veri setinde bir puanlama sistemi ile bir sıralamanın belirlenmesi söz konusudur ve hesaplanan puanlar Tablo 6.2'de gösterilmiştir. Tabloda özellik seçim işlemi ile elde edilen sonuçlara göre önem sırası, seçilen özelliğin adı ve hesaplanan puanlar gösterilmiştir. Tablo incelendiğinde en çok puana sahip olarak ilk sırada yer alan özelliğin bitki indisi MSI olduğu görülmektedir. Puan sıralaması dikkate alındığında, ilk 25 özelliğin uydu görüntüsünün spektral bantlarının işlenmesiyle türetilen özellikler olduğu görülmektedir. Başka bir ifadeyle, dış kaynaktan edinilen özellikler seçilen ilk 25 özellik arasında yer almamaktadır. Ancak puan sıralamasına bakıldığında 26. sırada seçilen ilk dış kaynaklı yardımcı verinin ALOS-PALSAR uydu verisinden elde edilen yükseklik özelliği olduğu görülmektedir. Tabloda seçilen spektral bantların adları kısaltılarak verilmiştir. Belirlenen puanlama sıralamasına göre 7. sırada bulunan morfolojik operatörlerden genişleme (dilation) işlemi bulunmaktadır. Bu özellik mevcut Sentinel-2A uydu görüntüsünün kırmızı bandına uygulanmasıyla üretilmiş olan özelliktir. Puanlama sistemine göre 23. sırada seçilen spektral bant HSI renk uzay dönüşüm işlemlerinden biri olan yoğunluk (I), uydu görüntüsünün yakın kızılötesi kırmızı ve yeşil bantları ile oluşturulmuş kombinasyonuna uygulanmasıyla üretilen özelliği temsil etmektedir. Önemli sonuçlardan biri görüntü dönüşüm yöntemlerinden olan temel bileşenler analizi ile elde edilen ve %62,34 oranında bilgi taşıyan ilk bileşenin diğer bileşenlere göre daha az bir puan hesaplanarak tabloda son sıralarda yer almasıdır.

Tablo 6.2: Ki-kare özellik seçim yöntemiyle belirlenmiş veri seti.

Önem Sırası	Seçilen Özellik	Puanlama
1	MSI	1386,758
2	S (Kırmızı-Yeşil-Mavi)	847,197
3	CIg	825,911
4	BRBA	774,888
5	RDVI	696,808
6	H (Kırmızı-Yeşil-Mavi)	693,600
7	Kırmızı (Genişleme)	573,692
8	RENDVI	556,814
9	Kırmızı	531,465
10	S (NIR-Kırmızı-Yeşil)	529,358
11	SAVI	526,753
12	OSAVI	526,636
13	NDVI	526,579
14	IPVI	526,567
15	Kırmızı (Açılma)	489,034
16	VARI	470,868
17	Kırmızı (Kapanma)	465,428
18	NIR (Açılma)	434,845
19	I (Kırmızı-Yeşil-Mavi)	432,553
20	GNDVI	411,940
21	NIR	411,564
22	NIR (Aşınma)	402,496
23	I (NIR-Kırmızı-Yeşil)	396,302
24	Yeşil (Genişleme)	394,138
25	NIR (Genişleme)	387,993
26	Yükseklik	381,001
27	NIR (Kapanma)	376,886
28	Kırmızı (Aşınma)	347,210
29	Toprak türü	345,310
30	Yeşil	337,289
31	SWIR-2	336,790
32	Mavi (Genişleme)	333,662
33	Yeşil (Açılma)	332,356
34	SWIR-2 (Genişleme)	331,363
35	Nüfus	323,606
36	SWIR-2 (Kapanma)	322,107
37	SWIR-2 (Açılma)	298,035
38	H (NIR-Kırmızı-Yeşil)	278,185
39	SWIR-2 (Aşınma)	268,781
40	Yeşil (Kapanma)	259,290
41	Mavi	255,250
42	WRI	251,516
43	Mavi (Açılma)	248,901
44	Kırmızı kenar-1 (Genişleme)	248,445
45	BAEI	248,345
46	Kırmızı (Varyans)	247,597
47	Yeşil (Varyans)	236,380
48	DBSI	232,816
49	Kırmızı kenar-1 (Açılma)	229,898
50	Kırmızı kenar-1 (Kapanma)	229,040

Tablo 6.2: Devam.

Önem Sırası	Seçilen Özellik	Puanlama
51	Kırmızı kenar-1	228,590
52	GLI	221,479
53	Mavi (Kapanma)	220,392
54	Mavi (Varyans)	218,818
55	Kırmızı (Ortalama)	217,136
56	NIR (Varyans)	211,938
57	GUF	211,493
58	SWIR-1 (Genişleme)	208,745
59	Yeşil (Ortalama)	198,665
60	Kırmızı kenar-1 (Aşınma)	195,522
61	SWIR-1 (Kapanma)	191,819
62	SWIR-1	190,775
63	Yeşil (Aşınma)	186,216
64	SWIR-1 (Açılma)	182,526
65	TDVI	177,539
66	TB-2	168,606
67	NIR (Ortalama)	164,930
68	Mavi (Aşınma)	163,485
69	SWIR-1 (Varyans)	157,027
70	SWIR-1 (Aşınma)	155,059
71	Mavi (Ortalama)	151,622
72	SWIR-2 (Varyans)	131,891
73	Kırmızı kenar-2 (Varyans)	125,950
74	Kırmızı kenar-1 (Varyans)	124,074
75	SATVI	111,774
76	Daraltılmış NIR (Zıtlık)	107,779
77	Kırmızı kenar-3 (Zıtlık)	102,675
78	Kırmızı kenar-2 (Zıtlık)	99,823
79	Daraltılmış NIR (Farklılık)	94,855
80	Kırmızı kenar-3 (Farklılık)	92,196
81	SWIR-1 (Ortalama)	91,650
82	Kırmızı kenar-2 (Farklılık)	88,377
83	SWIR-2 (Ortalama)	84,371
84	Kırmızı kenar-3 (Varyans)	84,156
85	B8A Varyans	83,550
86	NDWI	81,150
87	NIR (Homojenlik)	80,853
88	Kırmızı kenar-1 (Ortalama)	79,498
89	Kırmızı kenar-2 (Genişleme)	76,243
90	Kırmızı kenar-2 (Ortalama)	76,159
91	Mavi (Zıtlık)	70,448
92	Toprak Sıcaklığı	68,994
93	TB-3	68,351
94	Mavi (Farklılık)	66,603
95	NIR (Farklılık)	64,356
96	Yağış Miktarı	59,300
97	Kırmızı kenar-2 (Kapanma)	59,292
98	Daraltılmış NIR (Ortalama)	55,598
99	Kırmızı kenar-3 (Genişleme)	55,112
100	TB-1	53,408

6.2.2. ReliefF Yöntemi ile Boyut İndirgeme

ReliefF özellik seçim yöntemi ile belirlenen veri setlerinin RO algoritması ile sınıflandırılması sonucu genel doğruluklar hesaplanmıştır. Bir puanlama sistemine göre belirlenen boyut indirgeme uygulamasından 10, 20, 30, 40, 50, 75, 100 ve 135 banttan oluşan sekiz veri seti oluşturulmuştur. Veri setlerinin RO algoritması kullanılarak sınıflandırılması ile en yüksek doğruluk performansına 50 banttan oluşan veri seti ile (VS-5) ulaşılmıştır. 50 banttan oluşan bu veri setinde bulunan özelliklerin sıralamaları Tablo 6.3'te gösterilmiştir. Tabloda önem sırası, seçilen spektral özelliğin adı ve puan sıralaması yer almaktadır. Tablo incelendiğinde en fazla puan, HSI renk uzay dönüşümünün uydu görüntüsünün kırmızı, yeşil ve mavi bantlarından oluşan RGB bant kombinasyonuna uygulanmasıyla üretilen ton (H) özelliği olduğu sonucuna ulaşılmıştır. Sıralamaya göre ilk on bant arasında yer alan SAVI, OSAVI, NDVI ve IPVI spektral indisleri için hesaplanan puanların aynı olduğu görülmektedir. Bu sonuca göre söz konusu dört spektral indis için sınıflandırma işlemlerinde aynı katkıyı sağlayacağını söylemek mümkündür. Ayrıca puan sıralama sonuçları incelendiğinde Ki-kare yönteminden farklı olarak, seçilen ilk on bant arasında dış kaynaklı yardımcı veri olan toprak türü özelliğinin olduğu sonucuna ulaşılmıştır. Tabloda özellik seçim işlemine göre ilk 50 bant içerisinde yer alan diğer dış kaynaklı yardımcı veriler yağış haritası, toprak sıcaklığı haritası ve nüfus haritaları olduğu görülmektedir. Ayrıca görüntü dönüşüm yöntemlerinden olan temel bileşenler analizi ile elde edilen ve %7,28 oranında bilgi taşıyan üçüncü temel bileşenin (TB-3) seçilen özellikler arasında mevcut olduğu ancak %62,34 oranında bilgi taşıyan ilk bileşenin ReliefF tekniğiyle seçilmediği sonucu ortaya çıkmıştır.

Tablo 6.3: ReliefF özellik seçim yöntemiyle belirlenmiş veri seti.

Önem Sırası	Seçilen Özellik	Puanlama
1	H (Kırmızı-Yeşil-Mavi)	0,278
2	H (NIR-Kırmızı-Yeşil)	0,249
3	GNDVI	0,233
4	RENDVI	0,221
5	SAVI	0,220
6	OSAVI	0,220
7	NDVI	0,220
8	IPVI	0,220
9	Toprak Türü	0,209
10	S (NIR-Kırmızı-Yeşil)	0,197
11	RDVI	0,195
12	I (NIR-Kırmızı-Yeşil)	0,176
13	I (Kırmızı-Yeşil-Mavi)	0,175
14	VARI	0,172
15	NIR (Açılma)	0,165
16	MSI	0,164
17	TDVI	0,162
18	BRBA	0,157
19	NIR (Aşınma)	0,156
20	Yakın kızılötesi	0,154
21	CIg	0,152
22	DBSI	0,152
23	NIR (Kapanma)	0,151
24	NIR (Genişleme)	0,145
25	S (Kırmızı-Yeşil-Mavi)	0,140
26	Kırmızı (Genişleme)	0,130
27	Kırmızı (Kapanma)	0,115
28	Kırmızı	0,115
29	SWIR-2 (Genişleme)	0,010
30	SATVI	0,098
31	SWIR-2	0,098
32	Kırmızı (Açılma)	0,098
33	TB-3	0,096
34	SWIR-2 (Kapanma)	0,096
35	Yağış	0,094
36	Yeşil (Genişleme)	0,093
37	GLI	0,092
38	BAEI	0,088
39	SWIR-1 (Genişleme)	0,087
40	SWIR-2 (Açılma)	0,084
41	Toprak Sıcaklığı	0,082
42	Kırmızı kenar (Genişleme)	0,081
43	SWIR-1 (Kapanma)	0,080
44	Yeşil	0,080
45	SWIR-1	0,080
46	Yeşil (Kapanma)	0,079
47	WRI	0,078
48	Nüfus	0,075
49	Kırmızı kenar-1 (Kapanma)	0,075
50	SWIR-1 (Açılma)	0,075

6.2.3. Bilgi Kazancı Yöntemi ile Boyut İndirgeme

Bilgi Kazancı özellik seçim yöntemi, diğer yöntemlerde olduğu gibi yardımcı veriler ve uydu görüntüsünün spektral bantlarından oluşan 170 özelliğe sahip yüksek boyutlu veriye uygulanmıştır. Bu işlem sonucunda puan değerlerine göre bir önem sırası belirlenmiştir. Diğer yöntemlerde olduğu gibi elde edilen önem sırası göz önüne alınarak 10, 20, 30, 40, 50, 75, 100 ve 135 özellikten oluşan sekiz veri seti belirlenmiştir. Bu veri setleri ile RO algoritması kullanılarak sınıflandırma işlemleri gerçekleştirilmiştir. Sonuçlar analiz edildiğinde en yüksek sınıflandırma doğruluğuna 100 banttan oluşan (VS-7) veri setinin kullanılmasıyla elde edildiği sonucuna ortaya çıkmıştır. 100 banttan oluşan bu veri setinde hangi spektral bantların bulunduğu Tablo 6.4'te gösterilmiştir. Tablo incelendiğinde en çok puana sahip olarak ilk sırada yer alan bandın bitki indisi GNDVI olduğu görülmektedir. Puan sıralaması dikkate alındığında, ilk 46 özelliğin uydu görüntüsünün bantları kullanılarak türetilen özellikler olduğu ve dış kaynaklı yardımcı verilerin seçilen ilk 46 özellik arasında bulunmadığı görülmektedir. Ancak puan sıralamasına göre 47. sırada ilk seçilen dış kaynaklı yardımcı verinin toprak sıcaklığı olduğu görülmektedir. Tabloda en fazla puan HSI renk uzay dönüşüm işleminin uydu görüntüsünün RGB bant kombinasyonuna uygulanmasıyla üretilen ton (H) özelliği olduğu sonucu ortaya çıkmıştır. Tabloda özellik seçim işlemine göre ilk 50 bant içerisinde yer alan dış kaynaklı yardımcı verilerin yağış miktarı ve toprak sıcaklığını temsil eden bantlar olduğu gözlemlenmektedir. Ayrıca puan sıralamasına göre belirlenen ilk 100 özellik arasında dış kaynaklı yardımcı veriler olan yükseklik, nüfus, toprak türü ve küresel kentsel ayak izi özelliklerinin seçildiği de görülmektedir. Tez çalışması kapsamında kullanılan sekiz dış kaynaklı verinin altısı en yüksek sınıflandırma performansının sağlandığı veri setinde bulunduğu sonucu ortaya çıkmıştır.

Tablo 6.4: InfG özellik seçim yöntemiyle belirlenmiş veri seti.

Önem Sırası	Seçilen Özellik	Puanlama
1	GNDVI	1,325
2	CIg	1,325
3	BRBA	1,325
4	Kırmızı (Genişleme)	1,236
5	Kırmızı (Kapanma)	1,233
6	RENDVI	1,231
7	Kırmızı	1,222
8	VARI	1,201
9	TB-1	1,198
10	H (Kırmızı-Yeşil-Mavi)	1,192
11	Kırmızı (Açılma)	1,169
12	Kırmızı kenar-1 (Kapanma)	1,168
13	Kırmızı kenar-1	1,165
14	Kırmızı kenar-1 (Genişleme)	1,165
15	BAEI	1,163
16	Yeşil (Genişleme)	1,162
17	Yeşil (Kapanma)	1,159
18	SWIR-2 (Genişleme)	1,153
19	WRI	1,146
20	SWIR-2 (Kapanma)	1,141
21	SATVI	1,140
22	MSI	1,136
23	TDVI	1,133
24	OSAVI	1,133
25	SAVI	1,133
26	NDVI	1,133
27	IPVI	1,132
28	SWIR-2	1,131
29	Yeşil	1,130
30	RDVI	1,124
31	Kırmızı kenar-1 (Açılma)	1,115
32	SWIR-2 (Açılma)	1,099
33	Mavi (Kapanma)	1,098
34	Yağış miktarı	1,095
35	H (NIR-Kırmızı-Yeşil)	1,077
36	Mavi	1,077
37	I (Kırmızı-Yeşil-Mavi)	1,077
38	Mavi (Genişleme)	1,075
39	Yeşil (Açılma)	1,071
40	SWIR-1	1,063
41	EVI	1,062
42	SWIR-1 (Genişleme)	1,060
43	DBSI	1,055
44	SWIR-1 (Kapanma)	1,049
45	Kırmızı (Aşınma)	1,025
46	SWIR-1 (Açılma)	1,019
47	Toprak Sıcaklığı	1,015
48	Yeşil (Açılma)	0,998
49	S (NIR-Kırmızı-Yeşil)	0,997
50	SWIR-2 (Aşınma)	0,993

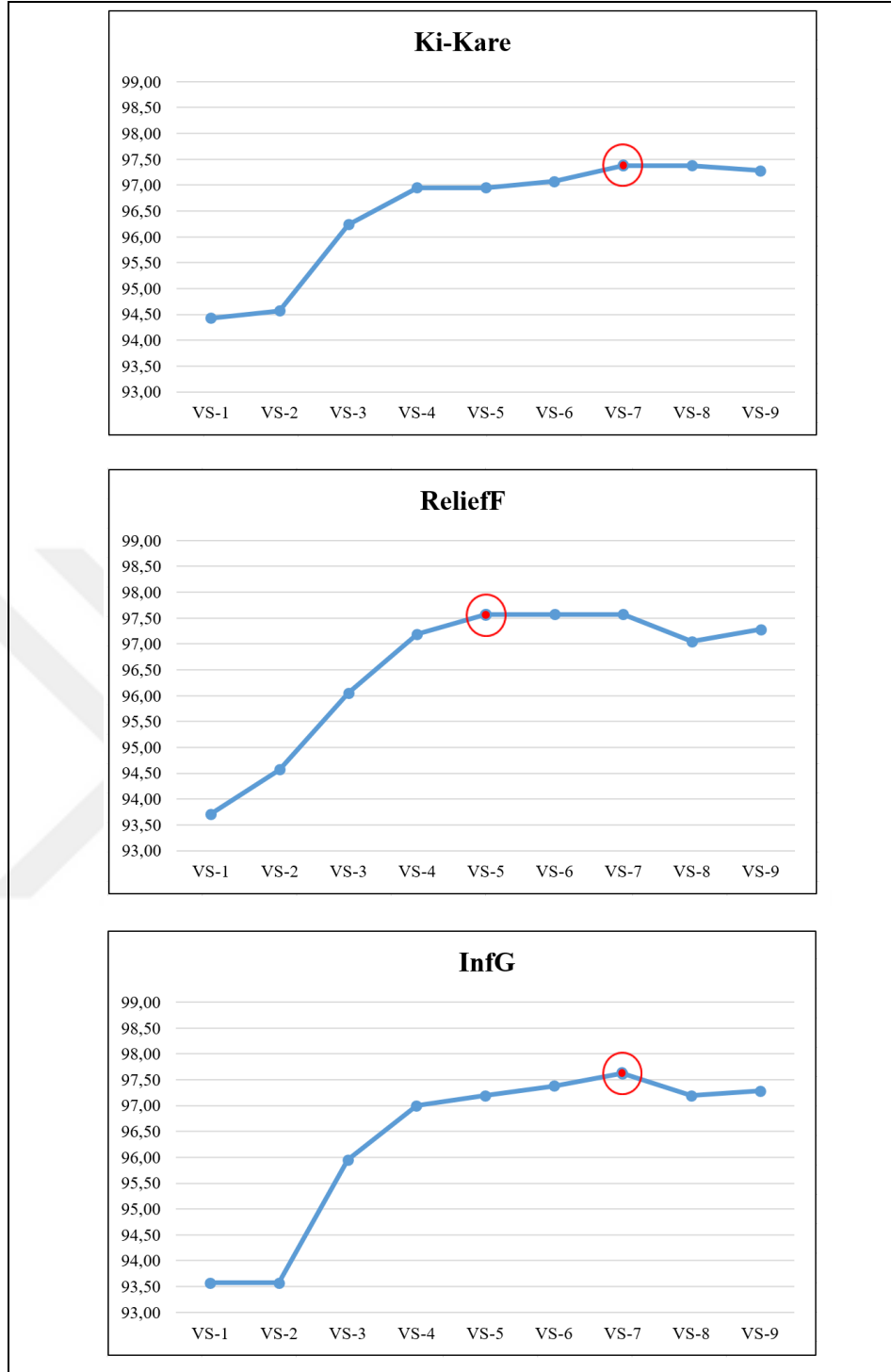
Tablo 6.4: Devam.

Önem Sırası	Seçilen Özellik	Puanlama
51	Kırmızı kenar-1 (Aşınma)	0,973
52	GARI	0,951
53	TB-3	0,930
54	Yeşil (Aşınma)	0,915
55	SWIR-1 (Aşınma)	0,914
56	NIR (Açılma)	0,874
57	Mavi (Aşınma)	0,862
58	NIR	0,841
59	GLI	0,827
60	NIR (Aşınma)	0,826
61	Kırmızı kenar-2 (Genişleme)	0,825
62	NIR (Kapanma)	0,794
63	Kırmızı kenar-2 (Kapanma)	0,781
64	S (Kırmızı-Yeşil-Mavi)	0,775
65	Yükseklik	0,764
66	I (NIR-Kırmızı-Yeşil)	0,761
67	NIR (Genişleme)	0,751
68	TB-2	0,748
69	GEMI	0,722
70	Daraltılmış NIR (Genişleme)	0,719
71	Kırmızı kenar-2	0,713
72	Kırmızı kenar-3 (Genişleme)	0,705
73	Daraltılmış NIR (Kapanma)	0,686
74	NDWI	0,685
75	Kırmızı kenar-3 (Kapanma)	0,675
76	Daraltılmış NIR	0,671
77	Kırmızı kenar-2 (Açılma)	0,663
78	Kırmızı kenar-3	0,637
79	Daraltılmış NIR (Açılma)	0,631
80	Kırmızı kenar-3 (Açılma)	0,601
81	Nüfus	0,564
82	Kırmızı kenar-2 (Aşınma)	0,545
83	Daraltılmış NIR (Aşınma)	0,535
84	Toprak Türü	0,526
85	Kırmızı kenar-3 (Aşınma)	0,525
86	NIR (Ortalama)	0,251
87	Kırmızı (Ortalama)	0,241
88	Kırmızı (Varyans)	0,239
89	SWIR-2 (Varyans)	0,236
90	GUF	0,230
91	SWIR-2 (Ortalama)	0,228
92	Yeşil (Varyans)	0,227
93	Kırmızı kenar-1 (Ortalama)	0,223
94	NIR (Varyans)	0,222
95	Yeşil (Ortalama)	0,221
96	Kırmızı kenar-1 (Varyans)	0,214
97	SWIR-1 (Ortalama)	0,189
98	Kırmızı kenar-3 (Farklılık)	0,186
99	SWIR-1 (Varyans)	0,184
100	Kırmızı kenar-2 (Zıtlık)	0,182

6.2.4. Sonuçların Karşılaştırılması

Ki-kare, ReliefF ve InfG ile belirlenen veri setleri RO algoritması ile sınıflandırılarak genel doğruluklar hesaplanmıştır. Sağlıklı bir değerlendirmenin yapılabilmesi için bütün veri setlerinde, RO sınıflandırma algoritmasına ait parametreler aynı değerde kullanılmıştır. Ormanda bulunan ağaç sayısının temsil edildiği 'n_estimators' parametresi 100 olarak işlemler yapılmıştır. ve diğer hiper parametreler ise scikit-learn kütüphanenin varsayılan değerleri seçilerek işlemler yapılmıştır. Şekil 6.3'te üç özellik seçim yöntemi ile belirlenen veri setlerine ait sınıflandırma doğruluk performansları verilmiştir. Kırmızı daireler ile belirtilen yerler en yüksek doğruluğu sağlayan veri setlerini temsil etmektedir. Ki-kare ve InfG yöntemlerinde en yüksek doğruluk ilk 100 bandı içeren (VS-7) veri seti kullanılarak elde edilirken, ReliefF algoritması ile en yüksek doğruluğa ilk 50 bandın temsil edildiği (VS-5) veri seti ile ulaşılmıştır.

Dokuz veri setine ayrı ayrı uygulanan RO algoritması ile Ki-kare yöntemi için hesaplanan en yüksek doğruluk %97,38 (VS-7), ReliefF metodu için hesaplanan en yüksek doğruluk %98,24 (VS-5) ve InfG için hesaplanan en yüksek doğruluk %98,14 (VS-7) olarak belirlenmiştir. Ayrıca RO algoritması uygulanarak tespit edilen en yüksek doğruluğa sahip bu üç veri seti; RO, XGB ve DVM algoritmasının yanı sıra söz konusu sınıflandırıcıların ortak olarak kullanıldığı oylama sınıflandırma tekniği ile değerlendirilmiştir. Hesaplanan sınıflandırma sonuçları Tablo 6.5'te gösterilmiştir. Ki-kare ve InfG özellik seçim yöntemine RO algoritması uygulanarak belirlenen en veri setleri için oylama tekniğiyle en yüksek sınıflandırma doğrulukları hesaplanmıştır. ReliefF özellik seçim yöntemi için ise XGB makine öğrenme algoritması ve oylama tekniğiyle en yüksek genel doğruluk hesaplanmıştır.

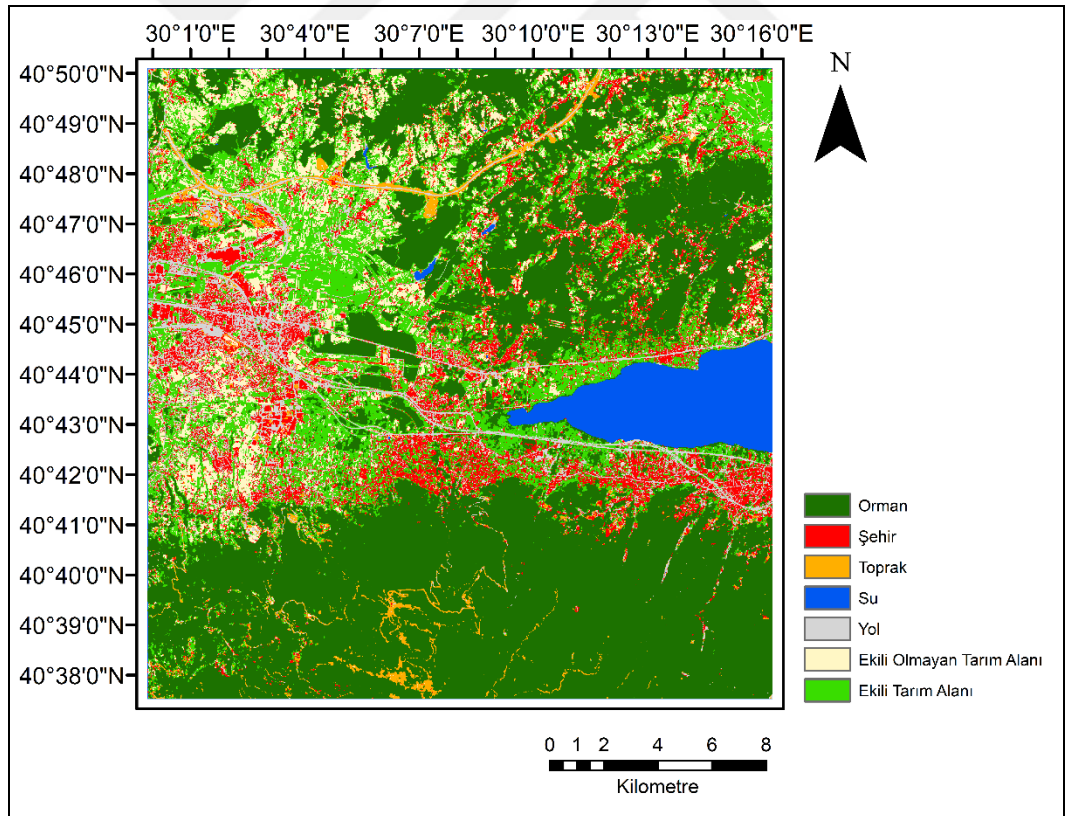


Şekil 6.3: Filtre tabanlı yöntemlerle belirlenen veri setlerine RO algoritmasının uygulanmasıyla hesaplanan genel doğruluk grafikleri.

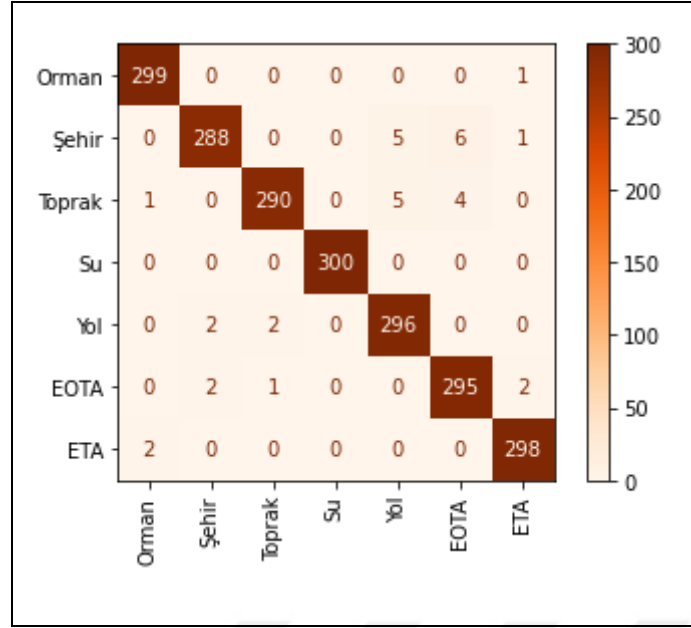
Tablo 6.5: Filtre tabanlı yöntemlerle belirlenen veri setleri için hesaplanan sınıflandırma doğrulukları.

Algoritma	Ki-kare	ReliefF	InfG
RO	97,38	97,57	97,62
XGB	98,29	98,24	98,00
DVM	97,00	96,81	97,07
Oylama	98,38	98,24	98,14

Farklı sınıflandırıcıların RO algoritması ile belirlenen veri setlerine uygulanmasındaki amaç sınıflandırma doğruluk analizlerinin daha detaylı bir şekilde yapılmasının hedeflenmesidir. Filtre tabanlı yöntemlerle belirlenen veri setleri için hesaplanan en yüksek doğruluk performansına (% 98,38) ait tematik harita Şekil 6.4'te gösterilmiştir. Ayrıca bu sınıflandırma işlemi sonucuna ait hata matrisi Şekil 6.5 ile belirtilmiştir.



6.4: Ki-kare metoduyla belirlenen VS-7'nin oylama tekniğiyle üretilen tematik haritası.



6.5: Ki-kare metoduyla belirlenen VS-7'nin oylama tekniğiyle sınıflandırılması sonucu üretilen hata matrisi.

Hata matrisinde, satırlarda yer alan değerler referans piksellerinin sayısını gösterirken, sütunlardaki değerler ise tahmin edilen piksel sayılarını temsil edilmektedir. Hata matrisi incelendiğinde şehir sınıfına ait pikseller diğer sınıflara göre daha düşük sınıflandırma performansı ile sınıflandırıldığı sonucuna ulaşılmaktadır. Ayrıca şehir sınıfına ait olan piksellerin, en fazla yol ve ekili olmayan tarım alanı sınıflarına ait olan pikseller ile karıştığı görülmektedir.

6.3. Sarmal Tabanlı Boyut İndirgeme

Genetik algoritma (GA), ardışık ileri yönde seçim (AİS) ve özyinelemeli özellik seçim yöntemleri (ÖÖS) birer sarmal tabanlı özellik seçim yaklaşımlarıdır. Bu yöntemler bir değerlendirme fonksiyonu (fitness function) ile birlikte kullanılmaktadır. Tez kapsamında bu üç yöntem rastgele orman sınıflandırıcısı ile değerlendirilmiş ve veri setleri tespit edilmiştir. Söz konusu üç sarmal özellik seçim yaklaşımı 170 bantlı yüksek boyutlu görüntüye uygulanarak 10, 20, 30, 40, 50, 75, 100 ve 135 bant içeren sekiz veri seti elde edilmiştir. Veri setleri VS-1, VS-2, VS-3, VS-4, VS-5, VS-6, VS-7, VS-8 ve 170 özellikten oluşan veri seti VS-9 olarak temsil edilmektedir. Tüm özellik seçim işlemleri Python 3.8.3 programlama dilinde,

Anaconda'nın Jupyter Notebook ara yüzü kullanılarak gerçekleştirilmiş ve analiz edilmiştir. Sarmal tabanlı özellik seçim yaklaşımları ile boyut indirgeme uygulamalarında, diğer yöntemlere göre çok daha fazla zaman maliyetinin olduğu sonucuna varılmıştır. Ortak sınıflandırıcı olarak diğer tüm yöntemlerde olduğu gibi RO algoritması kullanılmış ve veri setleri için genel doğruluklar hesaplanmıştır.

GA ile RO sınıflandırıcısı kullanılarak en yüksek doğruluğun hesaplandığı veri setinin belirlenmesi için gereken zaman maliyeti 32 dakika 11 saniye, AİS ve RO sınıflandırıcısı için zaman maliyeti 1 saat 11 dakika 31 saniye son olarak ÖÖS ve RO sınıflandırıcısı için zaman maliyeti 1 dakika 1 saniye olmuştur. Bu nedenle sarmal tabanlı özellik seçim yaklaşımlarında diğer yöntemlere göre daha fazla zaman maliyetinin olduğu sonucuna ulaşılmıştır. Zaman maliyetinin çok fazla olması, GA yaklaşımında yinelemeli olarak en uygun alt kümeyi tespit etmeyi hedefleyen çalışma prensibinden kaynaklanmaktadır. AİS özellik seçim metodunda yinelemeli olarak değerlendirilen kümeye özellikler eklenerek en yüksek doğruluğa ulaşma isteği zaman maliyetinin artmasına sebep olmaktadır. ÖÖS tekniğinde ise özellikler özyinelemeli bir şekilde ortadan kaldırılarak kalan özellikler ile bir model oluşturulması söz konusu olduğu için diğer sarmal yaklaşımlara göre daha az zaman maliyeti gerektirmektedir.

6.3.1. Genetik Algoritma

GA sarmal tabanlı yaklaşımda değerlendirme fonksiyonu olarak RO algoritmasının kullanılmasıyla sekiz ve 170 özellikten oluşan veri setleri elde edilmiştir. Bu veri setleri RO sınıflandırıcısı ile analiz edilmiş ve belirlenen en yüksek doğruluk 50 banttan oluşan (VS-5) veri setine ait olduğu sonucuna ulaşılmıştır. Sarmal tabanlı özellik seçim yöntemlerinde filtre veya gömülü yöntemlerde olduğu gibi bir puanlama sistemi söz konusu değildir. Sarmal yöntemlerde bir özellik kümesi belirlenmesi ilkesi mevcuttur. Başka bir ifadeyle, özellik seçim işlemi ile puan sıralaması yerine özelliklerin bulunduğu bir alt küme elde edilmektedir. Uygulamada GA ile RO sınıflandırıcısının kullanıldığı sarmal yaklaşım için popülasyon sayısı 100, mutasyon parametresi 0,1 ve çaprazlama parametresi 0,5 olarak seçilmiş ve özellik seçim işlemleri yapılmıştır. Tablo 6.6'de bu yaklaşımla belirlenen ve 50 banttan oluşan alt küme ifade edilmiştir.

Tablo 6.6: GA ve RO uygunluk fonksiyonuyla belirlenen küme.

Özellik Türü	Seçilen Özellik
Dış Kaynaklı Veriler	Eğim
	Yağış Miktarı
	Nüfus
	GUF
	Yükseklik
HSI Dönüşümü	H (Kırmızı-Yeşil-Mavi)
	H (NIR-Kırmızı-Yeşil)
	S (Kırmızı-Yeşil-Mavi)
	S (NIR-Kırmızı-Yeşil)
	I (Kırmızı-Yeşil-Mavi)
Spektral İndisler	BAEI
	GEMI
	GLI
	NDTI
	NDWI
	VARI
Temel Bileşenler	TB-1
	TB-3
Morfolojik İşlemler	Mavi (Genişleme)
	Mavi (Aşınma)
	Mavi (Açılma)
	Yeşil (Açılma)
	Kırmızı Kenar-1 (Aşınma)
	Kırmızı Kenar-1 (Kapanma)
	Kırmızı Kenar-3 (Genişleme)
	NIR (Aşınma)
	Daraltılmış NIR (Aşınma)
	Daraltılmış NIR (Açılma)
	SWIR-1 (Açılma)
	SWIR-2 (Kapanma)
	Mavi (Entropi)
Doku Özellikleri	Mavi (Korelasyon)
	Mavi (Ortalama)
	Mavi (Zıtlık)
	Yeşil (Entropi)
	Yeşil (Farklılık)
	Yeşil (Korelasyon)
	Kırmızı (Açısal ikinci moment)
	Kırmızı (Varyans)
	Kırmızı Kenar- 2 (Açısal İkinci Moment)
	Kırmızı Kenar-3 (Farklılık)
	Kırmızı Kenar-3 (Homojenlik)
	Kırmızı Kenar-3 (Zıtlık)
	NIR (Açısal ikinci moment)
	Daraltılmış NIR (Homojenlik)
	Daraltılmış NIR (Korelasyon)
	Daraltılmış NIR (Ortalama)
	SWIR-2 (Entropi)
	SWIR-2 (Varyans)
Spektral Bant	SWIR-1

6.3.2. Ardışık İleri Yönde Özellik Seçim Yaklaşımı

Ardışık ileri yönde seçim (AİS) sarmal tabanlı yaklaşım ve değerlendirme fonksiyonu olarak RO algoritmasının kullanılmasıyla diğer uygulamalarda olduğu gibi alt kümeler oluşturulmuş ve analiz edilmiştir. Ayrıca bu veri setleri RO sınıflandırıcısı ile analiz edilmiş ve sınıflandırma doğrulukları hesaplanmıştır. Tüm bantları içeren (170 bant) yüksek boyutlu verinin sınıflandırılması ile hesaplanan doğruluğun diğer sekiz veri seti için hesaplananlardan daha fazla olduğu sonucuna ulaşılmıştır. Sarmal yöntemlerin karşılaştırıldığı Şekil 6.6’da gösterildiği üzere, AİS-RO sarmal tabanlı yaklaşımda oluşturulan veri setleri diğer sarmal yöntemlerde olduğu gibi daha yüksek bir sınıflandırma performansı sağlamamıştır. Bunun sebebi yöntemin yinelemeli olarak çalışma ilkesinden kaynaklanmaktadır. AİS algoritmasında her yinelemede en yüksek katkıyı sağlayacak bandın sınıflandırma işlemine dâhil edilmesi söz konusu olduğu için oluşturulan alt kümelerle yapılan sınıflandırma işlemi için uygun bir yöntem olmadığı sonucuna ulaşılmıştır.

6.3.3. Özyinelemeli Özellik Seçim Yaklaşımı

Bir başka sarmal tabanlı özellik seçim yaklaşımı olan özyinelemeli özellik seçim yönteminin (ÖÖS) RO algoritmasının değerlendirme fonksiyonu olarak kullanılmasıyla diğer uygulamalarda olduğu gibi veri setleri üretilmiş ve analizleri yapılmıştır. Söz konusu veri setlerinin sınıflandırma performansları RO sınıflandırıcısı ile gerçekleştirilmiştir. Belirlenen en yüksek genel doğruluk 50 özellikten oluşan (VS-5) veri setine ait olduğu sonucu ortaya çıkmıştır. ÖÖS ve uygunluk fonksiyonu olarak RO algoritmasının değerlendirilmesi sarmal tabanlı bir özellik seçim yaklaşımı olduğundan puanlama sistemi yerine bir özellik kümesi belirlenmesi doğruluk ilkesi mevcuttur. Bu nedenle filtre ve gömülü özellik seçim işlemlerinde olduğu gibi önem sırası yerine bir özellik kümesi belirlenir. Tablo 6.7’de en yüksek genel doğruluk ile belirlenen ve 50 banttan oluşan bu küme verilmiştir. Diğer özellik seçim yöntemlerinden farklı olarak herhangi bir doku özelliğinin belirlenen kümede yer almaması önemli bir sonuç olarak ortaya çıkmıştır.

Tablo 6.7: ÖÖS ve RO uygunluk fonksiyonuyla belirlenen küme.

Özellik Türü	Seçilen Özellik
Dış Kaynaklı Veriler	Yükseklik
	Eğim
	Yağış Miktarı
HSI Dönüşümleri	H (Kırmızı-Yeşil-Mavi)
	H (NIR- Kırmızı-Yeşil)
	S (Kırmızı-Yeşil-Mavi)
	S (NIR-Kırmızı-Yeşil)
	I (Kırmızı-Yeşil-Mavi)
	I (NIR-Kırmızı-Yeşil)
Spektral İndisler	BRBA
	CIg
	DBSI
	GARI
	GNDVI
	MSI
	IPVI
	NDTI
	NDVI
	NDWI
	OSAVI
	RDVI
	RENDVI
	SAVI
	TDVI
	VARI
	WRI
Temel Bileşen	TB-3
Morfolojik İşlemler	Mavi (Genişleme)
	Mavi (Açılma)
	Mavi (Kapanma)
	Yeşil (Genişleme)
	Yeşil (Aşınma)
	Yeşil (Açılma)
	Yeşil (Kapanma)
	Kırmızı (Genişleme)
	Kırmızı (Aşınma)
	Kırmızı (Açılma)
	Kırmızı (Kapanma)
	Kırmızı kenar-2 (Genişleme)
	Kırmızı kenar-2 (Kapanma)
	Kırmızı kenar-3 (Genişleme)
	Kırmızı kenar-3 (Kapanma)
	NIR (Aşınma)
	NIR (Açılma)
	Daraltılmış NIR (Aşınma)
	Daraltılmış NIR (Açılma)
	SWIR-2 (Genişleme)
Spektral Bantlar	Kırmızı
	Kırmızı kenar-2
	NIR

6.3.4. Sonuçların Karşılaştırılması

GA, AİS ve ÖÖS sarmal özellik seçim yöntemleri ve uygunluk fonksiyonu RO algoritmasının kullanılmasıyla çalışmanın amacına uygun olarak her sarmal yöntemde dokuz veri seti oluşturulmuştur ve bu veri setlerinin genel doğrulukları hesaplanmıştır. Gerçekçi bir değerlendirme yapabilmek adına uygulanan RO sınıflandırıcısı için ormandaki ağaç sayısının belirlendiği parametre hepsinde aynı sayıda seçilmiştir. Şekil 6.6’da veri setlerine ait sınıflandırma doğrulukları verilmiştir. Kırmızı daireler ile belirtilen yerler sınıflandırma analizleri sonucunda en yüksek doğruluğu sağlayan veri setlerini temsil etmektedir. GA ve ÖÖS sarmal yöntemlerde en yüksek doğruluk 50 bant ile sağlanırken AİS sarmal yöntemi etkisiz kalmıştır. AİS yönteminde belirlenen veri setlerinin hiçbiri 170 banttan oluşan tüm verinin sınıflandırılmasından daha yüksek bir doğruluk sağlamamıştır.

Veri setlerinin analiz edilmesiyle, GA yöntemi için elde edilen en yüksek doğruluk % 98,00 (VS-5), ÖÖS metodu için elde edilen en yüksek doğruluk % 97,81 (VS-5) olarak hesaplanmıştır. AİS yönteminde oluşturulan alt kümelerin sınıflandırma performansları, 170 banttan oluşan yüksek boyutlu verinin performansından daha kötü olduğu tespit edilmiştir. Ayrıca GA ile ÖÖS sarmal özellik seçim yöntemleri uygulanarak tespit edilen en yüksek doğruluğa sahip bu iki veri seti, DVM, XGB ve RO algoritmasının ortak olarak kullanıldığı oylama yöntemi ile sınıflandırılmıştır. Sonuçlar Tablo 6.9’da yüzde olarak grafikler halinde ifade edilmiştir. XGB sınıflandırıcısında öğrencinin türü olan ve ağaç topluluğunu oluşturan parametre ‘gbtree’ olarak işlemler yapılmıştır. GA, AİS ve ÖÖS sarmal yaklaşımlı özellik seçimleri için en yüksek doğruluklar oylama sınıflandırma yöntemi ile hesaplandığı sonucuna varılmıştır. Hesaplanan en yüksek sınıflandırma doğruluklarına ait performanslar Tablo 6.8’de belirtilmiştir.

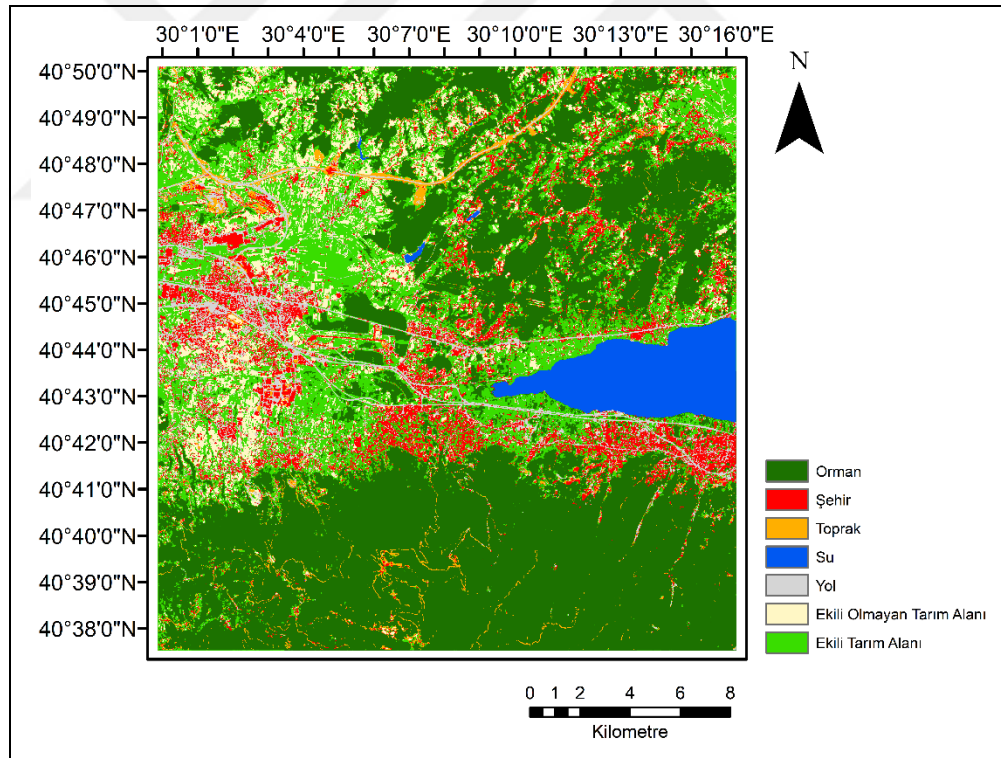


Şekil 6.3: Sarmal yaklaşımlarla belirlenen veri setlerine RO algoritmasının uygulanmasıyla hesaplanan genel doğruluk grafikleri.

Tablo 6.8: Sarmal yaklaşımlarla belirlenen veri setleri için hesaplanan sınıflandırma doğrulukları.

	GA	AİS	ÖÖS
RO	98,00	97,29	97,81
XGB	98,19	97,00	98,14
DVM	97,33	97,54	97,57
Oylama	98,29	97,90	98,23

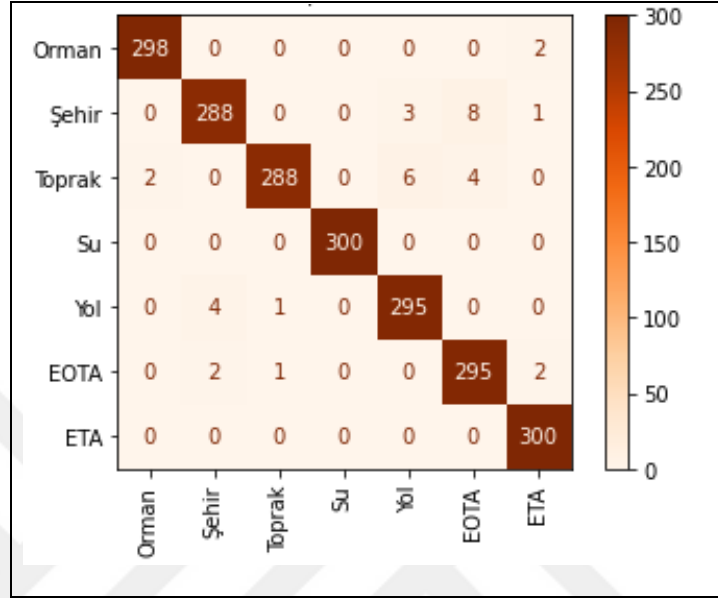
RO algoritması ile belirlenen veri setlerine çeşitli sınıflandırma algoritmalarının uygulanmasındaki amaç doğruluk analizlerini detaylandırmaktır. Sarmal tabanlı yaklaşımlar ile belirlenen veri setleri için hesaplanan en yüksek doğruluk performansına ait (% 98,29) tematik harita Şekil 6.7’de gösterilmiştir. Ayrıca bu sınıflandırma işlemi sonucuna ait hesaplanan hata matrisi Şekil 6.8 ile gösterilmiştir.



Şekil 6.7: GA ve RO uygunluk fonksiyonuyla belirlenen VS-5’nin oylama tekniğiyle sınıflandırılması sonucu üretilen tematik haritası.

Hata matrisinde satırlar referans piksellerini, sütunlar ise tahmin edilen pikselleri temsil etmektedir. Hata matrisi incelendiğinde toprak ve şehir sınıfına ait pikseller

diğer sınıflara göre daha düşük sınıflandırma performansı gösterdiği sonucuna ulaşılmaktadır. Ayrıca toprak sınıfına ait olan pikseller en fazla yol sınıfına ait olan pikseller ile karıştığı, şehir sınıfına ait olan pikseller ise ekili olmayan tarım alanları pikselleri ile karıştığı görülmektedir.



Şekil 6.8: GA ve RO uygunluk fonksiyonuyla belirlenen VS-5'in oylama tekniğiyle sınıflandırılması sonucu üretilen hata matrisi.

6.4. Gömülü Boyut İndirgeme

Gömülü özellik seçimi yöntemlerinin analizinin yapılabilmesi için, LightGBM, HSIC Lasso ve RO olmak üzere 3 farklı yöntem 170 banttan oluşan yüksek boyutlu görüntüye uygulanarak 10, 20, 30, 40, 50, 75, 100 ve 135 bant içeren sekiz veri seti elde edilmiştir. Veri setleri VS-1, VS-2, VS-3, VS-4, VS-5, VS-6, VS-7, VS-8 ve tüm özelliklerin olduğu veri seti VS-9 olarak temsil edilmektedir. Özellik seçim işlemlerinin tamamı Python 3.8.3 programlama dilinde ve Anaconda'nın Jupyter Notebook ara yüzü kullanılarak analiz edilmiştir. LightGBM yöntemi kullanılarak gerçekleştirilen özellik seçim işlemi 1,14 saniye, son zamanlarda geliştirilmiş HSIC Lasso algoritması uygulanarak yapılan özellik seçim işlemi 2,5 milisaniye ve RO algoritması ile yapılan özellik seçim işlemi 2,93 saniye sürmüştür. Zaman maliyetlerinden anlaşılabileceği üzere gömülü özellik seçim yöntemleri ile çok hızlı bir şekilde özellikler için önem sırası belirlenmektedir. Yöntemlerin zaman maliyetinin çok düşük olmasının yanında, diğerlerine göre en hızlı şekilde sonuç hesaplayan

yöntem HSIC Lasso olmuştur. Özellik seçimi sınıflandırma modeli ile gerçekleştirildiği için, sarmal özellik seçim yaklaşımlarının aksine daha az hesaplama maliyetine ihtiyaç duydukları sonucuna ulaşılmaktadır.

6.4.1. LightGBM ile Boyut İndirgeme

LightGBM özellik seçim yöntemi ile oluşturulan 9 veri seti RO algoritması ile analiz edildikten sonra belirlenen en yüksek doğruluğun VS-2 ile sağlandığı sonucuna ulaşılmıştır. 20 banttan oluşan bu veri setinde hangi özelliklerin bulunduğu Tablo 6.9'da gösterilmiştir. Çalışmada LightGBM algoritmasının eğitim aşamasında ağaç sayısı 100 ve artırma tipi geleneksel gradyan artırma karar ağacı (gbdt) olarak belirlenmiştir. Söz konusu özellik seçim yöntemi uygulanarak filtre tabanlı yöntemlerde olduğu gibi bir puanlama sistemi ile önem sırası belirlenmiş ve daha sonra RO algoritması ile sınıflandırma işlemi gerçekleştirilmiştir. Tabloda ilk sütun seçilen spektral bandın önem sırasını, ikinci sütun özelliğin adını ve son sütun hesaplanan puan değerini göstermektedir. Tablo incelendiğinde en çok puana sahip olarak ilk sırada yer alan özellik ALOS-PALSAR uydusundan temin edinilen yani bir dış kaynaklı veri olan yükseklik özelliği olmuştur. Daha önce filtre tabanlı yöntemler ile belirlenen önem sıralamaları incelendiğinde ilk sırada uydu görüntüsünün bantlarından üretilen yardımcı verilerin olduğu ifade edilmiştir. LightGBM gömülü özellik seçim yönteminde puanlama sistemine göre belirlenen önem sırasında ilk sırada dış kaynaklı veri olan yükseklik bandının bulunması önemli bir sonuç olarak ortaya çıkmıştır. Bir başka önemli bulgu ise belirlenen veri setinde, tez çalışmasında kullanılan sekiz dış kaynaklı özelliğin beşinin seçilmiş olmasıdır. Ayrıca tüm işlemler kıyaslandığında en az veri setinin belirlendiği yöntem LightGBM olduğu sonucuna ulaşılmıştır. Bu yöntemde en yüksek sınıflandırma performansına belirlenen 20 bant ile ulaşılmıştır.

Tablo 6.9: LightGBM özellik seçim yöntemiyle belirlenmiş veri seti.

Önem Sırası	Seçilen Özellik	Puanlama
1	Yükseklik	601
2	H (Kırmızı-Yeşil-Mavi)	589
3	NDTI	439
4	S (Kırmızı-Yeşil-Mavi)	344
5	H (NIR-Kırmızı-Yeşil)	334
6	VARI	282
7	Daraltılmış NIR (Aşınma)	282
8	Yeşil (Genişleme)	251
9	Kırmızı (Genişleme)	223
10	S (NIR-Kırmızı-Yeşil)	213
11	TB-3	210
12	Kırmızı kenar-2 (Genişleme)	201
13	Nüfus	196
14	Eğim	179
15	Yağış miktarı	174
16	Daraltılmış NIR (Genişleme)	164
17	GUF	151
18	TB-1	141
19	BAEI	131
20	Yeşil (Açılma)	126

6.4.2. Hilbert-Schmidt Serbestlik Kriteri Lasso ile Boyut İndirgeme

Hilbert-Schmidt Serbestlik Kriteri Lasso (HSIC Lasso) özellik seçim yöntemi ile belirlenen veri setlerinin RO algoritması ile analiz edilmesiyle belirlenen en yüksek sınıflandırma doğruluğuna VS-4 ile ulaşılmıştır. 40 banttan oluşan bu veri setinde bulunan özelliklerin isimleri Tablo 6.12’de verilmiştir. Diğer gömülü özellik seçim işlemlerinde olduğu gibi, HSIC Lasso özellik seçim yönteminde de puanlama sistemi ile bir önem sırası elde edilmesi söz konusudur. Bu yöntemde bir cezalandırma sistemi mevcuttur. Bu nedenle çalışmanın amacına uygun olması sebebiyle, elenen özellikler için kendi aralarında tekrar değerlendirmek suretiyle özelliklere puan atanması sağlanmıştır. Tablo incelendiğinde filtre tabanlı Ki-kare yönteminde olduğu gibi en çok puana sahip olarak ilk sırada yer alan özellik bitki indisi MSI olmuştur. Ayrıca LightGBM yönteminde olduğu gibi ALOS-PALSAR uydusundan sağlanan yükseklik verisi üst sıralarda (5. sıra) yer alarak sınıflandırma işlemine katkı sağlayan bir özellik olduğu pekiştirilmiştir.

Tablo 6.10: HSIC Lasso özellik seçim yöntemiyle belirlenmiş veri seti.

Önem Sırası	Seçilen Özellik	Puanlama
1	MSI	1,000
2	CIg	0,883
3	H (Kırmızı-Yeşil-Mavi)	0,747
4	TB-3	0,641
5	Yükseklik	0,546
6	Kırmızı kenar-3	0,475
7	GUF	0,434
8	Toprak Türü	0,396
9	EVI	0,342
10	H (NIR-Kırmızı-Yeşil)	0,336
11	Nüfus	0,321
12	BAEI	0,292
13	NDWI	0,253
14	Daraltılmış NIR (Aşınma)	0,245
15	Kırmızı (Genişleme)	0,205
16	Toprak Sıcaklığı	0,194
17	Kırmızı kenar-3 (Ortalama)	0,186
18	Kırmızı kenar-3 (Genişleme)	0,183
19	Yaşıl (Açılma)	0,133
20	Mavi (Farklılık)	0,133
21	Yağış	0,126
22	NIR (Zıtlık)	0,115
23	Mavi (Varyans)	0,106
24	S (NIR-Kırmızı-Yeşil)	0,103
25	Kırmızı kenar-3 (Zıtlık)	0,091
26	Kırmızı kenar-1 (Zıtlık)	0,091
27	Kırmızı (Zıtlık)	0,090
28	Kırmızı kenar-3 (Zıtlık)	0,080
29	SWIR-1 (Zıtlık)	0,079
30	Kırmızı kenar-2 (Korelasyon)	0,075
31	Kırmızı kenar-3 (Açılma)	0,074
32	SWIR-1 (Açısal ikinci moment)	0,059
33	Yeşil (Zıtlık)	0,057
34	NDTI	0,055
35	SWIR-2 (Genişleme)	0,051
36	SWIR-2 (Ortalama)	0,037
37	NIR (Varyans)	0,032
38	SWIR-2 (Zıtlık)	0,019
39	Mavi (Açısal ikinci moment)	0,014
40	Kırmızı kenar-2 (Ortalama)	0,011

6.4.3. Rastgele Orman Algoritması ile Boyut İndirgeme

Rastgele Orman (RO) özellik seçim yöntemi bir gömülü boyut indirgeme yöntemidir. Tez kapsamında sınıflandırma ve boyut indirgeme işlemlerinde RO algoritmasının tercih edilmesinin sebebi birçok farklı alanda kabul görmüş bir yöntem olmasından kaynaklanmaktadır. Algoritma yardımcı verilerin eklenmesiyle 170 banttan oluşan yüksek boyutlu veri setine, önem sıralarının belirlendiği boyut indirgeme işleminin gerçekleştirilmesi için uygulanmıştır. Eğitim aşamasında diğer uygulamalarda olduğu gibi ağaç sayısı 100 olarak işlem yapılmıştır. İşlemin sonucunda 10, 20, 30, 40, 50, 75, 100 ve 135 özellikten oluşan veri setleri belirlenmiştir. Belirlenen veri setleri RO algoritması ile analiz edildiğinde en yüksek doğruluk 100 özellikten oluşan (VS-7) veri setinin kullanılmasıyla hesaplandığı sonucu ortaya çıkmıştır. VS-7’de hangi bantların bulunduğu Tablo 6.11’de gösterilmiştir. Tabloda ilk sütun seçilen spektral bandın önem sırasını, ikinci sütun seçilen özelliğin adını ve son sütun hesaplanan puanların değerlerini göstermektedir. Tablo incelendiğinde en çok puana sahip olarak ilk sırada yer alan bant bitki indisi Cİg olduğu görülmektedir. Seçilen ilk on özelliğin çoğu bant oranlamaları ile üretilen spektral indisler olmuştur. Bunun sebebi çalışma alanında vejetasyon yoğunluğunun fazla olması olarak değerlendirilmiştir. Elde edilen önemli bir bulgu LightGBM ve HSIC Lasso gömülü boyut indirgeme yöntemlerinde olduğu gibi, RO boyut indirgeme yönteminde de yükseklik özelliğinin en yüksek performansın elde edildiği veri setinde üst sıralarda yer almış olmasıdır. Bu sonuçtan yola çıkarak dış kaynaklı yardımcı veri olan yükseklik özelliğinin özellikle topoğrafik farklılığın fazla olduğu arazi yüzeylerinde, sınıflandırma işlemine olumlu etki ederek, daha yüksek doğrulukta haritaların üretilmesine imkân sağladığı söylenebilir. Elde edilen önemli bir bulgu ise yükseklik haricinde üst sıralarda herhangi bir dış kaynaklı yardımcı verinin bulunmamasıdır. Tablo incelendiğinde belirlenen özellikler arasında yükseklik verisinden sonra seçilen ilk dış kaynaklı verinin eğim olduğu görülmektedir.

Tablo 6.11: RO özellik seçim yöntemiyle belirlenmiş veri seti.

Önem Sırası	Seçilen Özellik	Puanlama
1	CIg	0,0306
2	NDTI	0,0298
3	Kırmızı (Genişleme)	0,0228
4	GNDVI	0,0226
5	Yeşil (Açılma)	0,0224
6	RDVI	0,0221
7	DBSI	0,0219
8	BAEI	0,0211
9	NIR	0,0211
10	TB-3	0,0211
11	RENDVI	0,0205
12	S (Kırmızı-Yeşil-Blue)	0,0203
13	NIR (Açılma)	0,0194
14	NDVI	0,0176
15	H (NIR-Kırmızı-Yeşil)	0,0175
16	Kırmızı	0,0174
17	Yükseklik	0,0173
18	BRBA	0,0167
19	MSI	0,0164
20	VARI	0,0163
21	Kırmızı kenar-3 (Kapanma)	0,0156
22	H (Kırmızı-Yeşil-Mavi)	0,0154
23	SWIR-2 (Genişleme)	0,0151
24	Yeşil (Aşınma)	0,0146
25	I (NIR-Kırmızı-Yeşil)	0,0147
26	NIR (Genişleme)	0,0145
27	NIR (Kapanma)	0,0144
28	Kırmızı kenar-2 (Genişleme)	0,0143
29	WRI	0,0132
30	Kırmızı (Açılma)	0,0128
31	Mavi (Kapanma)	0,0128
32	Kırmızı (Aşınma)	0,0125
33	Mavi	0,0116
34	Kırmızı (Kapanma)	0,0115
35	SAVI	0,0113
36	TB-2	0,0110
37	Mavi (Genişleme)	0,0107
38	Yeşil (Genişleme)	0,0107
39	NIR (Aşınma)	0,0106
40	IPVI	0,0101
41	Yeşil	0,0101
42	Kırmızı kenar-2	0,0101
43	Eğim	0,0101
44	TDVI	0,0098
45	Kırmızı kenar-2 (Kapanma)	0,0097
46	GLI	0,0095
47	OSAVI	0,0094
48	Yeşil (Kapanma)	0,0090
49	Kırmızı kenar-1	0,0086
50	Daraltılmış NIR (Aşınma)	0,0084

Tablo 6.11: Devam.

Önem Sırası	Seçilen Özellik	Puanlama
51	Kırmızı kenar-3	0,0082
52	Kırmızı kenar-3 (Genişleme)	0,0081
53	Daraltılmış NIR (Genişleme)	0,0080
54	GARI	0,0078
55	S (NIR-Kırmızı-Yeşil)	0,0075
56	I (Kırmızı-Yeşil-Mavi)	0,0075
57	SWIR-2 (Kapanma)	0,0074
58	SWIR-2	0,0074
59	GUF	0,0072
60	Mavi (Açılma)	0,0068
61	Daraltılmış NIR (Açılma)	0,0068
62	Kırmızı kenar-1 (Genişleme)	0,0067
63	Yağış	0,0066
64	SWIR-2 (Açılma)	0,0065
65	EVI	0,0063
66	TB-1	0,0063
67	SATVI	0,0057
68	GEMI	0,0053
69	SWIR-1 (Genişleme)	0,0052
70	SWIR-1	0,0052
71	Kırmızı kenar-1 (Kapanma)	0,0051
72	Mavi (Kapanma)	0,0050
73	NDWI	0,0047
74	Kırmızı kenar-1 (Açılma)	0,0041
75	Kırmızı kenar-3 (Aşınma)	0,0040
76	Daraltılmış NIR (Kapanma)	0,0037
77	Daraltılmış NIR	0,0036
78	SWIR-1 (Kapanma)	0,0035
79	SIPI	0,0034
80	Toprak Türü	0,0034
81	Kırmızı kenar-2 (Açılma)	0,0032
82	Nüfus	0,0031
83	SWIR-1 (Açılma)	0,0030
84	Kırmızı kenar-3 (Açılma)	0,0025
85	Toprak Sıcaklığı	0,0024
86	Kırmızı kenar-1 (Aşınma)	0,0020
87	Kırmızı kenar-2 (Aşınma)	0,0020
88	SWIR-2 (Aşınma)	0,0013
89	SWIR-1 (Aşınma)	0,0009
90	Kırmızı (Ortalama)	0,0009
91	NIR (Korelasyon)	0,0008
92	Kırmızı kenar-1 (Varyans)	0,0007
93	Kırmızı kenar-2 (Açısal moment)	0,0007
94	Bakı	0,0007
95	Kırmızı kenar-1 (Ortalama)	0,0007
96	Kırmızı kenar-3 (Farklılık)	0,0006
97	Mavi (Varyans)	0,0006
98	Kırmızı kenar-1 (Korelasyon)	0,0006
99	NIR (Varyans)	0,0006
100	Kırmızı kenar-2 (Farklılık)	0,0005

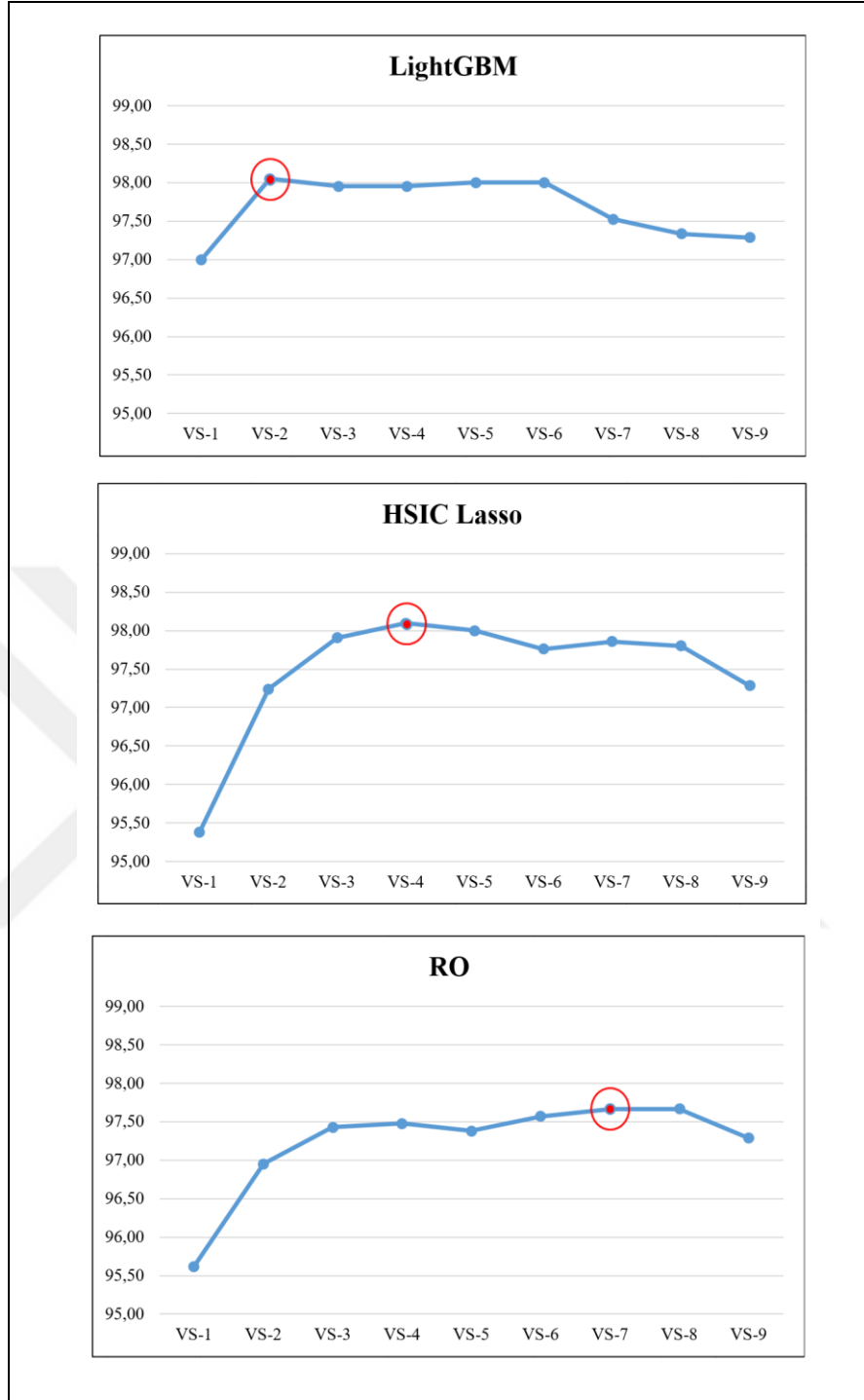
6.4.4. Sonuçların Karşılaştırılması

LightGBM, HSIC Lasso ve RO gömülü özellik seçim yöntemleri ile belirlenen veri setleri RO algoritması ile analiz edilmiş ve sonuçlar karşılaştırılmıştır. Doğru bir değerlendirme yapabilmek için RO sınıflandırma algoritması bütün veri setlerine aynı parametreler ile uygulanmıştır. Şekil 6.9’da üç özellik seçim yöntemi ile belirlenen bütün veri setleri için hesaplanmış sınıflandırma doğruluklarına ait grafikler verilmiştir. Kırmızı daireler ile belirtilen yerler en yüksek doğruluğu sağlayan veri setlerini göstermektedir. LightGBM özellik seçim yöntemi ile belirlenen veri setleri, RO algoritması ile sınıflandırılmış en yüksek doğruluk (%98,05) VS-2 ile sağlanmıştır. HSIC Lasso özellik seçim yöntemi ile tespit edilen dokuz veri seti, RO sınıflandırıcı ile analiz edilmiş ve en yüksek doğruluk (%98,10) VS-4 ile elde edilmiştir. Son olarak RO (%97,67) ile elde edilen dokuz veri setine yine RO algoritması kullanılarak sınıflandırma işlemi gerçekleştirilmiş ve en yüksek doğruluk VS-7 ile elde edilmiştir.

Ayrıca RO algoritması uygulanarak tespit edilen en yüksek doğruluğa sahip bu üç veri seti; DVM, XGB ve RO algoritmasının ortak olarak kullanıldığı oylama sınıflandırma yöntemi ile sınıflandırılmıştır. Sonuçlar tablo 6.12’de verilmiştir. Belirlenen veri setlerinin analiz edilmesi sonucu en yüksek performanslar oylama yöntemi ile elde edilmiştir. Ayrıca kullanılan gömülü özellik seçim yöntemleri arasında en iyi sınıflandırma sonucu HSIC Lasso özellik seçim yöntemi ile belirlenen VS-4 ile hesaplanmıştır.

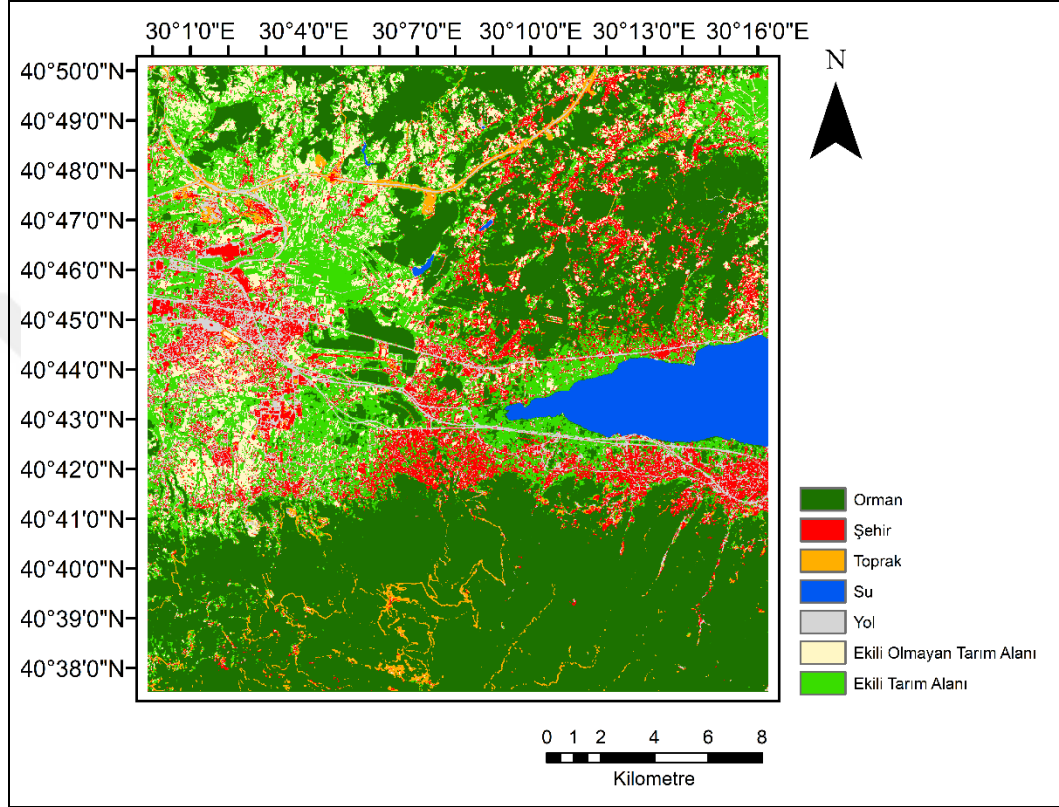
Tablo 6.12: Gömülü yöntemlerle belirlenen veri setleri için hesaplanan sınıflandırma doğrulukları.

	LightGBM	HSIC Lasso	RO
RO	98,05	98,10	97,67
XGB	98,14	98,43	98,19
DVM	97,03	95,24	96,10
Oylama	98,29	98,48	98,33



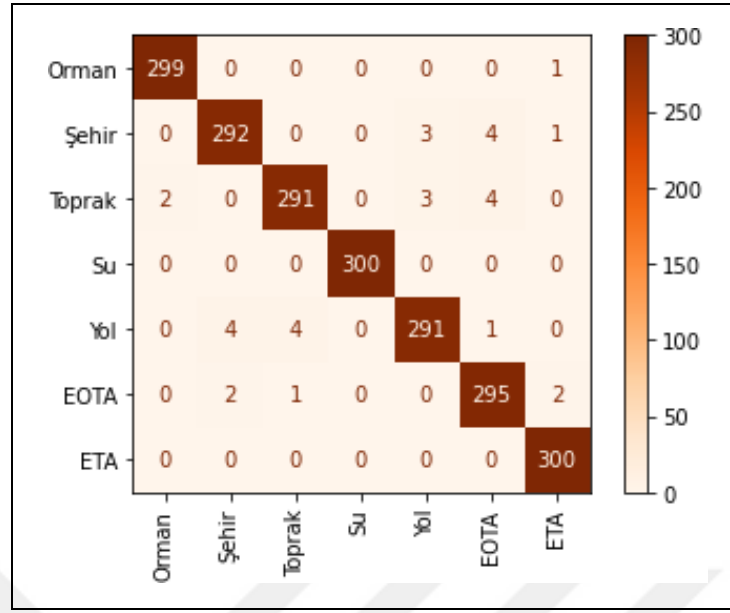
Şekil 6.9: Gömülü yöntemlerle belirlenen veri setlerine RO algoritmasının uygulanmasıyla hesaplanan genel doğruluk grafikleri.

Farklı sınıflandırma algoritmalarının uygulanmasıyla doğruluk sonuçlarına ilişkin analizlerin detaylı bir şekilde yapılmıştır. Gömülü yöntemlerle belirlenen veri setleri için hesaplanan en yüksek doğruluk performansı ile hesaplanan (%98,48) ürün harita Şekil 6.10'da gösterilmiştir. Ayrıca bu sınıflandırma işlemi sonucuna ait hesaplanan hata matrisi Şekil 6.11 ile gösterilmiştir.



Şekil 6.10: HSIC Lasso metoduyla belirlenen VS-4'ün oylama tekniğiyle sınıflandırılması sonucu üretilen tematik harita.

Hata matrisinde satırlar toplamı referans piksellerini diğer kısım ise tahmin edilen pikselleri göstermektedir. Hata matrisi incelendiğinde toprak ve yol sınıfına ait pikseller diğer sınıflara göre daha düşük sınıflandırma performansı ile hesaplandığı sonucu ortaya çıkmıştır. Toprak pikselleri ekili olmayan tarım alan sınıfına ait pikseller ile karıştığı görülmektedir. Yol sınıfına ait pikseller ise daha çok şehir ve toprak sınıfı ile karıştığı görülmektedir.



Şekil 6.11: HSIC Lasso metoduyla belirlenen VS-4'ün oylama tekniğiyle sınıflandırılması sonucu üretilen hata matrisi.

6.5. McNemar's Testi ile Sınıflandırma Sonuçlarının Analizi

Filtre tabanlı, sarmal tabanlı ve gömülü özellik seçim yöntemlerine dayanan 9 algoritmanın kullanılmasıyla belirlenen veri setleri; RO, XGB ve DVM algoritmaları ile sınıflandırma işlemi gerçekleştirilmiştir. Ayrıca bu üç sınıflandırma yönteminin birlikte kullanılarak elde edilen ve daha güçlü bir sınıflandırıcı yöntemi olan oylamama tekniği, belirlenen veri setlerinin sınıflandırılmasında kullanılmıştır. Oylama tekniği, temel kolektif bir öğrenme metodudur. Dokuz farklı özellik seçim yöntemiyle belirlenen veri setleri RO algoritmasıyla sınıflandırılmış ve en yüksek sınıflandırma doğruluğu Ki-kare (VS-7), GA (VS-5), ve HSIC Lasso (VS-4) ile elde edilmiştir. Elde edilen bu üç veri seti RO, XGB ve DVM algoritmalarının beraber kullanıldığı oylama yöntemiyle sınıflandırılmıştır. Bununla beraber, hesaplanan hata matrisleri hesaplanmıştır. Genel doğrulukların değerlendirilip aralarındaki farkların istatistiksel olarak test edilmesi için Ki-kare dağılımının esas alındığı 2x2 boyutlu bir matris ile hesaplanan McNemar's (χ^2) testi uygulanmıştır. Ki-kare tablo dağılımının esas alındığı %95 güven aralığı için tablo değeri 3,84'tür. Tablo 6.13'te görüldüğü üzere HSIC Lasso ve Ki-kare özellik seçim yöntemi için hesaplanan değer (0,05) 3,84 olan tablo değerinden küçüktür ve bu nedenle iki sonuç için aralarında anlamlı bir fark

yoktur yorumu yapılır. Ki-kare ile GA ve uygunluk fonksiyonu olarak RO yönteminin uygulanması sonucu aralarında hesaplanan değer (4,36) tablo değerinden büyüktür. Bu nedenle aralarında anlamlı bir fark olduğu yorumu yapılmaktadır. Ayrıca GA ve uygunluk fonksiyonu RO kullanılması ile HSIC Lasso yöntemine göre belirlenen veri setlerinin oylama sınıflandırma tekniğiyle hesaplanan sınıflandırma sonuçları McNemar's (χ^2) testi ile analiz edilmiştir. Söz konusu hesaplanan değer (8,65) tablo değerinden büyük olduğu için aralarında anlamlı bir fark oldukları sonucu ortaya çıkmaktadır.

Tablo 6.13: Yapılan sınıflandırma işlemlerinin McNemar's istatistiksel testi ile analiz sonuçları.

Algoritmalar	McNemar' Testi
Ki-kare ve GA	4,36
GA ve HSIC Lasso	8,65
Hsic Lasso ve Ki-kare	0,05

7. SONUÇLAR ve TARTIŞMA

Teknoloji alanındaki gelişmelerle birlikte uzaktan algılama biliminin önemi de gün geçtikçe artmaktadır. Uzaktan algılanmış veriler, özellikle geniş alanlarda yapılan çalışmalarda uygun maliyet avantajına sahiptir. Dünya yüzeyine ilişkin uygulamalarda uzaktan algılanmış görüntüler yardımıyla çeşitli amaçlar doğrultusunda kullanılabilecek tematik haritalar üretilmektedir. Bu tematik haritaların, üretim aşamasında hangi sınıflandırma yönteminin tercih edilmesi gerektiği veya hangi verilerin kullanılması gerektiği soruları uygulamalar için büyük önem arz etmektedir. Uzaktan algılama bilimi yaygın bir kullanım alanına sahip olduğundan söz konusu iki durum önemli araştırma alanları haline gelmektedir. Doğruluk analizleri yapılarak sınıflandırma işlemi sonucunda elde edilen ürünlerin hangi oranda ve ne kadar başarılı üretildiği belirlenebilmektedir. Ayrıca uzaktan algılanmış görüntüler, arazi örtüsü/arazi kullanımı harita üretiminde ana veri kaynağı olarak kullanılmaktadır. Bunun yanında sınıflandırma doğruluğunu artıracak yardımcı verilerin değerlendirilmesi, yüksek doğrulukta tematik harita elde etmek ve daha kullanışlı bir ürün ortaya çıkarmak için kullanılmaktadır.

Bu tez kapsamında yardımcı verilerin tematik harita üretimine olan etkilerinin detaylı analizleri yapılması amaçlanmıştır. Bu amaç doğrultusunda önce Sentinel-2A uydu görüntüsünün mevcut spektral bantları ile sınıflandırma işlemleri gerçekleştirilmiştir. Yardımcı verilerin sınıflandırma sürecinde etkilerini gözlemlemek için elde edilen yardımcı veriler uydu görüntüsünün spektral bantlarına eklenerek 170 banttan oluşan yüksek boyutlu bir veri seti oluşturulmuştur. Bu yardımcı veriler mevcut uydu görüntüsünün spektral bantları kullanılarak çeşitli işlemler sonucunda üretilen; doku özellikleri, temel bileşenler analizi ile tespit edilen bileşenler, ton-doygunluk-yoğunluk renk uzayı dönüşümü ile elde edilen özellikler, spektral indisler ve morfolojik filtrelerin uygulanmasıyla elde edilen özelliklerdir. Bu özelliklerin yanı sıra farklı bir kaynaktan temin edilen yardımcı veriler de tez çalışması kapsamında analiz edilmiştir. Dış kaynaklı yardımcı veriler çalışma alanına ait küresel kentsel ayak izi haritası, nüfus haritası, sayısal yükseklik modelinin işlenmesiyle üretilen eğim, bakı ve yükseklik haritaları, yeryüzüne ait toprak türü haritası, yağış miktarı ve toprak sıcaklığını temsil eden haritalardır. Özelliklerin her birine ait değerlendirmenin yapılabilmesi için filtre tabanlı, sarmal tabanlı yaklaşımlar ve

gömülü özellik seçim yöntemlerinden dokuz tanesi bu çalışma kapsamında değerlendirilmiştir. Değerlendirme aşamasında özellik seçim yöntemleri ile 10, 20, 30, 40, 50, 75, 100 ve 135 banttan oluşan veri setleri belirlenmiştir. Bu veri setleri VS-1, VS-2, VS-3, VS-4, VS-5, VS-6, VS-7 ve VS-8 olarak temsil edilmiştir. Ayrıca 170 özellikten oluşan tüm veri seti VS-9 olarak temsil edilmektedir.

Uzaktan algılanmış Sentinel-2A verisinin mevcut 10 m ve 20 m mekânsal çözünürlükte olan spektral bantları kullanılarak 12,5 m'ye yeniden örneklenmiştir. Yardımcı veriler de 12,5 m mekânsal çözünürlüğe sahip olduğu için bu işlem gerçekleştirilmiştir. Söz konusu Sentinel-2A spektral bantları ile yapılan sınıflandırma işlemleri sonucunda RO algoritması ile %94,86 genel doğruluk, XGB algoritmasıyla %94,33 genel doğruluk, DVM algoritması ile %93,86 genel doğruluk ve söz konusu sınıflandırıcılar kullanılarak yapılan oylama tekniği işlemi sonucunda %95,48 oranında genel doğruluklar hesaplanmıştır.

Yardımcı veriler ile oluşturulan yüksek boyutlu görüntüye filtre tabanlı, sarmal tabanlı yaklaşımlar ve gömülü özellik seçim yöntemleri uygulanmıştır. Kullanılan filtre tabanlı özellik seçim yöntemlerinden ilki olan Ki-kare tekniği ile her bir özelliğe ait bir önem sırası belirlenmiş ve bu önem sırasına göre oluşturulan veri setleri RO algoritmasıyla sınıflandırılarak en yüksek doğruluğa (%97,38) 100 banttan oluşan veri setiyle (VS-7) ulaşıldığı tespit edilmiştir. İkinci filtre tabanlı özellik seçim yöntemi olan ReliefF söz konusu yüksek boyuttaki veriye uygulanarak her banda ait önem sırası puanlama sonuçları ile belirlenmiştir. Bu önem sıralaması göz önüne alınarak alt kümeler elde edilmiş ve RO yöntemi ile sınıflandırılarak doğruluk analizleri yapılmıştır. En yüksek doğruluk (%97,57) 50 banttan oluşan veri seti (VS-5) kullanılarak elde edilmiştir. Son olarak InfG filtre tabanlı özellik seçim yöntemi yüksek boyutlu veriye uygulanmış ve her banda ait önem sırası tespit edilmiştir. Diğer uygulamalarda olduğu gibi belirlenen önem sıralamasına göre 9 veri seti elde edilmiş ve RO yöntemi ile sınıflandırılarak doğruluk analizleri yapılmıştır. En yüksek doğruluk performansı (%97,62) 100 banttan oluşan veri seti (VS-7) kullanılarak hesaplanmıştır. Özellik seçim işlemlerinde aynı sınıflandırma algoritması ve aynı parametler kullanılarak gerçekçi bir analiz yapılması hedeflenmiştir. Ayrıca en yüksek doğruluk sağlayan veri setleri, RO, XGB ve DVM algoritmalarının birlikte ele alındığı oylama sınıflandırma yöntemiyle sınıflandırma performansları karşılaştırılmıştır. Ki-kare özellik seçimi ile belirlenen VS-7 için oylama sınıflandırma sonucu %98,38 ve

InfG ile tespit edilen VS-7 için oylama sınıflandırma sonucu %98,14 olarak hesaplanmıştır. ReliefF özellik seçimi ile tespit edilen veri seti ile en yüksek doğruluk XGB algoritması ve oylama tekniğiyle aynı şekilde %98,23 olarak hesaplanmıştır. Ki-kare özellik seçimi ile belirlenen en iyi sınıflandırma doğruluğunun hesaplandığı veri setinde sırasıyla yükseklik, nüfus, GUF, toprak sıcaklığı ve yağış haritaları bulunduğu sonucuna ulaşılmıştır. Tez çalışmasında kullanılan sekiz dış kaynaklı veriden beş tanesi bu veri seti içinde bulunmaktadır ve dış kaynaklı verilerin %62,50'sinin seçildiği görülmektedir.

Sarmal tabanlı özellik seçim yaklaşımları bir uygunluk fonksiyonu ile değerlendirilmektedir ve çalışma kapsamında uygunluk fonksiyonu olarak RO algoritması kullanılmıştır. İlk sarmal tabanlı özellik seçim yaklaşım olan GA, uyarlama fonksiyonu RO algoritmasıyla kullanıldığında belirlenen alt kümeler sınıflandırılarak en yüksek doğruluğa (%98,00) 50 banttan oluşan veri setiyle (VS-5) ulaşılmıştır. Diğer sarmal tabanlı yaklaşım olan AİS, RO uygunluk fonksiyonu ile söz konusu yüksek boyuttaki veriye uygulanarak sınıflandırma doğrulukları hesaplanmıştır. Sınıflandırma doğrulukları göz önüne alınarak diğer yöntemlerde olduğu gibi 10, 20, 30, 40, 50, 75, 100 ve 135 banttan oluşan veri setleri elde edilmiştir. En yüksek doğruluk (%97,29) 170 banttan oluşan yüksek boyutlu veri ile elde edilmiştir. Bu durumdan yola çıkarak AİS sarmal tabanlı özellik seçim yaklaşımının en kötü performansa sahip olduğu ve boyut indirgeme işlemleri için etkili bir yöntem olmadığı sonucuna varılmıştır. Değerlendirmeye alınan son RO uygunluk fonksiyonu ile birlikte kullanılan sarmal yaklaşım ÖÖS, yüksek boyutlu veriye uygulanmış ve diğerlerinde olduğu gibi aynı sayıda bant içeren veri setleri elde edilmiştir. En yüksek genel doğruluk (%97,81) 50 banttan oluşan veri seti (VS-5) kullanılarak hesaplanmıştır. Bütün özellik seçim işleminde aynı sınıflandırma algoritması kullanılarak gerçekçi bir analiz yapılması hedeflenmiştir. Ayrıca söz konusu farklı sarmal yaklaşımlar ile en yüksek doğrulukla hesaplanan bu üç veri seti, RO, XGB ve DVM algoritmaları ve oylama sınıflandırma yöntemiyle sınıflandırılarak doğruluk performansları karşılaştırılmıştır. GA ile belirlenen VS-5 için oylama sınıflandırma sonucu %98,29 ve AİS ile belirlenen VS-5 için oylama sınıflandırma sonucu %98,23 olarak hesaplanmıştır. Ayrıca daha önce de belirtildiği gibi; AİS sarmal yaklaşımı ve RO uygunluk fonksiyonu ile belirlenen her bir veri seti için hesaplanan sınıflandırma doğruluğu, yüksek boyutlu 170 banttan oluşan verinin

hesaplanan sınıflandırma doğruluğundan daha düşüktür. Buna rağmen diğer sonuçlarla karşılaştırmak için yüksek boyutlu verinin tamamı aynı algoritmalarla sınıflandırılmış ve 170 banttan oluşan yüksek boyutlu veri ile elde edilen en yüksek doğruluk %97,90 olarak hesaplanmıştır. Genetik algoritma yaklaşımı ile belirlenen en yüksek sınıflandırma performansına sahip veri setinde dış kaynaklı veriler eğim, yağış miktarı, nüfus, GUF ve yükseklik haritaları mevcuttur. Tez çalışmasında kullanılan sekiz dış kaynaklı veriden beş tanesi veri seti içinde bulunmaktadır ve dış kaynaklı verilerin %62,50'sinin seçildiği görülmektedir. Ayrıca seçilen veri setinin %10'unu söz konusu dış kaynaklı veriler oluşturmaktadır.

Gömülü özellik seçim yöntemlerinde optimum özellik alt kümesinin seçilmesi sınıflandırıcı tarafından gerçekleştirilmektedir. Değerlendirmeye alınan gömülü özellik seçim yöntemlerinden ilki olan LightGBM metodu ile her bir banda ait bir önem sırası belirlenmiştir. Puanlama sistemine dayalı olarak elde edilen bu önem sırasına göre diğer yöntemlerde olduğu gibi veri setleri belirlenmiştir. Belirlenen bu veri setleri RO algoritmasıyla sınıflandırılarak en yüksek doğruluk (%98,05) 20 banttan oluşan veri setiyle (VS-2) hesaplanmıştır. Son yıllarda üretilmiş olan özellik seçim yöntemi HSIC Lasso bu tez kapsamında kullanılan ikinci gömülü özellik seçim algoritmasıdır. Söz konusu 170 banttan oluşan yüksek boyuttaki veriye uygulanarak her banda ait önem sırası, hesaplanan puan sonuçlarına göre belirlenmiştir. Yöntemde bir cezalandırma sistemi mevcuttur. Bu nedenle çalışmanın amacına uygun olması adına, elenen özellikler için kendi aralarında tekrar değerlendirme yapılarak özelliklere puanlama sistemine göre önem sırası belirlenmiştir. Bu önem sıralaması göz önüne alınarak diğer tüm yöntemlerde olduğu gibi alt kümeler elde edilmiş ve RO yöntemi ile sınıflandırılarak doğruluk analizleri yapılmıştır. En yüksek doğruluk (%98,10) 40 banttan oluşan veri seti (VS-4) kullanılarak hesaplanmıştır. Son olarak RO gömülü özellik seçim yöntemi söz konusu veriye uygulanmış ve her banda ait önem sırası tespit edilmiştir. Diğer tüm uygulamalarda olduğu gibi belirlenen önem sıralamasına göre sekiz veri seti elde edilmiş ve yine RO yöntemi ile sınıflandırılarak hesaplanan doğruluk değerleri belirlenmiştir. En yüksek doğruluk (%97,67) 100 banttan oluşan veri setinin (VS-7) sınıflandırılması ile hesaplandığı sonucuna ulaşılmıştır. Diğer boyut indirgeme yöntemleri ile belirlenen veri setlerinde olduğu gibi gömülü yöntemlerle belirlenen veri setleri de RO, XGB ve DVM algoritmalarının ve bunların birlikte ele alındığı oylama sınıflandırma yöntemiyle sınıflandırılmıştır.

LightGBM özellik seçim yöntemi ile belirlenen VS-2 için oylama sınıflandırma sonucu %98,29, HSIC Lasso özellik seçim yöntemi ile belirlenen VS-4 için oylama sınıflandırma sonucu %98,48 ve son olarak RO algoritmasıyla tespit edilen VS-7 için oylama sınıflandırma sonucu %98,33 olarak hesaplanmıştır. HSIC Lasso gömülü özellik seçim algoritması ile belirlenen en yüksek sınıflandırma doğruluğunun hesaplandığı veri setinde sırayla bulunan dış kaynaklı veriler yükseklik, GUF, toprak türü, nüfus, toprak sıcaklığı ve yağış miktarı özellikleridir. Tez çalışmasına konu olan sekiz dış kaynaklı veriden altı tanesi algoritma ile belirlenen veri setinde bulunmaktadır ve dış kaynaklı verilerin %75'inin belirlenen veri setinde bulunduğu görülmektedir. Ayrıca belirlenen veri setinin %15'ini söz konusu dış kaynaklı veriler oluşturmaktadır.

Tüm işlemler dikkate alındığında en yüksek genel doğruluk (%98,48) HSIC Lasso ile belirlenen veri setinin oylama yöntemiyle değerlendirilmesi sonucu hesaplanmıştır. Sentinel-2A verisinin spektral bantlarının oylama yöntemi kullanılarak sınıflandırılması sonucu hesaplanan genel doğruluk (%95,48) ile karşılaştırıldığında %3'ün üstünde bir iyileşmenin olduğu gözlemlenmiştir. Bu nedenle tez çalışması kapsamında kullanılan yardımcı verilerin, özellik seçim yöntemlerinin ve sınıflandırma algoritmalarının analiz edilmesi ile sınıflandırma doğruluğunda önemli bir artış olduğu kanıtlanmıştır.

Tez çalışması kapsamında, özellik seçim yöntemlerinin yüksek boyutlu veri setine uygulanmasıyla sınıflandırma işlemi için en yüksek performansı sağlayan en optimum özelliklerin belirlenmesi hedeflenmiştir. Bunun yanında özellik seçim işlemleri gerçekleştirilirken gerekli olan zaman maliyetleri de analistler için oldukça önemlidir. Uygulama kısmında da görüldüğü üzere özellik seçim işlemleri uygulanırken farklı zaman maliyetleri ortaya çıkmıştır. İlk dikkat çeken nokta sarmal tabanlı boyut indirgeme yaklaşımları için gereken sürelerin filtre ve gömülü özellik seçim yöntemlerine göre çok daha fazla olmasıdır. Filtre ve gömülü özellik seçim yöntemlerinde uygulanan özellik seçim işlemleri saniye mertebesinde zaman maliyetiyle sonuç hesaplarken, özellikle sarmal tabanlı GA ve AİS yöntemleri yarım saat ve bir saati bulan yüksek zaman maliyetleriyle optimum veri seti belirlenmiştir. Ayrıca düşük zaman maliyetine sahip gömülü özellik seçim yöntemleriyle üretilen veri setleri işlendiğinde daha yüksek sınıflandırma doğruluklarının hesaplandığı gözlemlenmiştir. Bu nedenle hem zaman maliyeti az hem de üretilen tematik

haritaların kalitesinin yüksek olması nedeniyle boyut indirgeme yöntemlerinde, filtreleme ve gömülü özellik seçim yöntemlerinin kullanılması önerilebilir. Filtre tabanlı ve gömülü yöntemler arasında bir kıyas yapılırsa, uygulanan 6 farklı algoritma (Ki-kare, ReliefF, InfG, LightGBM, Hsic lasso, RO) için hesap zaman maliyetleri saniye mertebesinde olup sarmal yöntemlere göre oldukça düşüktür. Gömülü özellik seçim yöntemleri ile eğitim aşamasında ihtiyaç duyulan zaman maliyetleri, filtre tabanlı yöntemler için gerekli zaman maliyetlerine göre küçük farklarla daha fazla olduğu görülmektedir. Bunun nedeni filtre tabanlı yöntemlerin sınıflandırıcıdan bağımsız olarak istatistik hesaplar ile sonuç hesaplaması ve gömülü özellik seçim yöntemlerinin bir sınıflandırıcı algoritmasıyla eğitim işleminin gerçekleştirilmesinden kaynaklanmaktadır. Özellik seçim işlemlerinde en önemli konulardan biri; makine öğrenme algoritmalarıyla veri üzerinde işlem yapılırken sağlıklı sonuçlar elde edebilmek için eğitim aşamasında aşırı öğrenme (overfitting) sorununa dikkat edilmesi gerektiği hususudur. Son olarak oylama sınıflandırma tekniğiyle hesaplanan sonuçlar McNemar's testi ile analiz edilerek aralarında nasıl bir ilişki olduğu ortaya konulmuştur. HSIC Lasso ve Ki-kare için hesaplanan değer tablo değerinden küçüktür ve bu iki sonuç için aralarında anlamlı bir fark olmadığını göstermektedir. Ki-kare ve GA yaklaşımı ve RO ile belirlenen veri setlerinin analizleri sonucunda hesaplanan doğruluk değerleri McNemar's testi kullanılarak analiz edildiğinde iki veri seti için aralarında anlamlı bir fark olduğu sonucuna ulaşılmıştır. Ayrıca GA yaklaşımı ve RO algoritması ile HSIC Lasso yöntemiyle belirlenen veri setlerine ait sınıflandırma sonuçları, McNemar's testi ile analiz edilmiştir. Hesaplanan değer tablo değerinden büyük olduğu için aralarında anlamlı bir fark olduğu sonucu ortaya çıkmıştır.

Gelecekte teknolojinin daha hızlı gelişmesi ile birlikte, uydulardan ve çeşitli kaynaklardan veri elde etme olanağı artacaktır. Ayrıca yüksek çözünürlükte ve çeşitli içeriğe sahip uzaktan algılanmış verilerin elde edilmesi de mümkün hale gelecektir. Söz konusu veriler birçok amaç için kullanılabilecek daha yüksek doğrulukta sınıflandırılmış tematik haritaların üretilmesine katkı sağlayacaktır.

KAYNAKLAR

- Acar U., Bayram B., (2009), "Building extraction with morphology", 4th International Conference on Recent Advances in Spaces Technologies, 33-38, İstanbul, Turkey, 11-13 June.
- Afshar M., Usefi H., (2021), "Dimensionality reduction using singular vectors", Scientific Reports, 11 (1), 3832.
- Anthony G., Ruther H., (2007), "Comparison of feature selection techniques for SVM Classification", 10th International Symposium on Physical Measurement and Signature in Remote Sensing, 1-6, Davos, Switzerland, 12-14 Marc.
- Archibald R., Fann G., (2007), "Feature selection and classification of hyperspectral images with support vector machines", IEEE Geoscience and Remote Sensing Letters, 4 (4), 674-677.
- Belgiu M., Drăgut L., (2016), "Random forest in remote sensing: A review of applications and future directions", ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24-31.
- Benediktsson A., Swain P. H., Ersoy O. K., (1993), "Conjugate-gradient neural networks in classification of very-high-dimensional remote sensing data", International Journal of Remote Sensing, 14 (15), 2883-2903.
- Birdsey R., Angeles-Perez G., Kurz W. A., (2013), "Approaches to monitoring changes in carbon stocks for REDD+", Carbon Management, 4 (5), 519-537.
- Budak H., (2018), "Özellik seçim yöntemleri ve yeni bir yaklaşım", Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 22, 21-31.
- Bouzekri S., Lasbet A. A., Lachehab A., (2015), "A new spectral index for extraction of built-up area using Landsat-8 data", Journal of the Indian Society of Remote Sensing, 43 (4), 867-873.
- Breiman L., (2001), "Random forests", Machine Learning, 45 (2), 5-32.
- Campbell, J. B., Wyne R. H., (2011), "Introduction to remote sensing", 5th Edition, The Guilford Press.
- Camps-Valls G., Tuia D., Gomez-Chova L., Jimenez S., Malo J., (2011), "Remote sensing image processing", In: A. C. Bovik, Editors, "Synthesis Lectures on Image, Video, and Multimedia Processing", CO: Morgan & Claypool Publishers.
- Cheng Q., Jing L., Panahi A., (2006), "Principal component analysis with optimum order sample correlation coefficient for image enhancement", International Journal of Remote Sensing, 27 (16), 3387-3401.

Chen T., Guestrin C., (2016), "XGBoost: A scalable tree boosting system", 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'16 ACM, 785-794, San Francisco, California, USA, 13-17 August.

Chernov V., Alander J., Bochko V., (2015), "Integer-based accurate conversion between RGB and HSV color spaces", Computers and Electrical Engineering, 46, 328-337.

Chuvieco E., (2009), "Fundamentals of satellite remote sensing", 2nd Edition, CRS Press, Taylor & Francis Group.

Colkesen I., Kavzoglu T., (2018), "Selection of optimal object features in object-based image analysis using filter-based algorithms", Journal of the Indian Society of Remote Sensing, 46 (8), 1233-1242.

Cortes C., Vapnik V., (1995), "Support-vector networks", Machine Learning, 20, 273-297.

Crippen R. E., (1990), "Calculating the vegetation index faster", Remote Sensing of Environment, 34 (1), 71-73.

Çölkesen İ., (2015), "Yüksek çözünürlüklü uydu görüntüleri kullanarak benzer spektral özelliklere sahip doğal nesnelerin ayırt edilmesine yönelik bir metodoloji geliştirme", Doktora Tezi, İstanbul Teknik Üniversitesi.

Çölkesen İ., Kavzoğlu T., Yomralıoğlu T., (2015), "Uzaktan algılanmış görüntülerde optimum bantların seçiminde destek vektör makinelerinin kullanımı", Türkiye Ulusal Fotogrametri ve Uzaktan Algılama Birliği VIII. Sempozyumu, Konya, İstanbul, 21-23 Mayıs.

Çölkesen İ., Yomralıoğlu T., (2014), "Arazi örtüsü ve kullanımının haritalanmasında Worldview-2 uydu görüntüsü ve yardımcı verilerin kullanımı", Harita Dergisi, 152 (2), 12-24.

Dash M., Liu H., (1997), "Feature selection for classification", Intelligent Data Analysis, 1 (3), 131-156.

Demarchi L., Kania A., Ciekowski W., Piórkowski H., Oświecimska-Piasko Z., Chormański J., (2020), "Recursive feature elimination and random forest classification of Natura 2000 Grasslands in lowland river valleys of Poland based on airborne hyperspectral and LiDAR data fusion" Remote Sensing, 12 (11), 1842.

Doğaner A., (2020), "Topluluk öğrenme yöntemleri ile renal hücreli Karsinom'un tahmin edilmesi", Doktora Tezi, İnönü Üniversitesi.

El-Asmar H., Hereher M. E., (2011), "Change detection of the coastal zone of the Nile Delta using remote sensing", Environmental Earth Sciences, 62 (4), 769-777.

Esch T., Heldens, W., Hinner A., Keil M., Marconcini M., Roth A., Zeidler J., Dech S., Strano E., (2017), "Breaking new ground in mapping human settlements from space

– the global urban footprint”, *ISPRS Journal of Photogrammetry and Remote Sensing*, 134, 30-42.

Fahsi A., Tsegaye T., Tadesse W., Coleman T., (200), “Incorporation of digital elevations models with Landsat-TM data to improve land cover classification accuracy”, *Forest Ecology and Management*, 128 (1-2), 57-64.

Foody G. M., (2004), “Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy”, *Photogrammetry Engineering and Remote Sensing*, 70 (5), 627-633.

Gao B. C., (1996), “NDWI-A normalized difference water index for remote sensing of vegetation liquid water from space”, *Remote Sensing of Environment*, 58 (3), 257-266.

Geng R., Jin S., Fu B., Wang B., (2020), “Object-based wetland classification using multi-feature combination of ultra-high spatial resolution multispectral images”, *Canadian Journal of Remote Sensing*, 46 (6), 784-802.

Genuer R., Poggi J., Tuleau-Malot C., Genuer R., Poggi J., (2010), “Variable selection using random forests”, 31 (14), 2225-2236.

Georganos S., Grippa T., Vanhuyse S., Lennert M., Shimoni M., Kalogirou S., Wolff E., (2018), “Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application”, *GIScience and Remote Sensing*, 55 (2), 221-242.

Georganos S., Grippa T., Vanhuyse S., Lennert M., Shimoni M., Wolff E., (2018), “Very high resolution object-based land use-land cover urban classification using extreme gradient boosting”, *IEEE Geoscience and Remote Sensing Letters*, 15 (4), 607-611.

Géron A., (2019). “Hands-on machine learning with scikit-learn, keras & tensorflow”, 2nd Edition, O’Reilly Media.

Ghimire B., Rogan J., Galiano V. R., Panday P., Neeti N., (2012), “An evaluating of bagging, boosting, and random forest for land-cover classification in Cape Cod, Massachusetts, USA”, *GIScience & Remote Sensing*, 49 (5), 623-643.

Gitelson A., Merzlyak M. N., (1994), “Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. Leaves. spectral features and relation to chlorophyll estimation”, *Journal of Plant Physiology*, 143 (3), 286-292.

Gitelson A. A., Gritz Y., Merzlyak M. N., (2003), “Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves”, *Journal of Plant Physiology*, 160 (3), 271-282.

Gitelson A. A., Kaufman Y. J., Merzlyak M. N., (1996), “Use of a green channel in

remote sensing of global vegetation from EOS-MODIS”, *Remote Sensing of Environment*, 58 (3), 289-298.

Gitelson A. A., Stark R., Grits U., Rundquist D., Kaufman Y., Derry D., (2002), “Vegetation and soil lines in visible spectral space: A concept and technique for remote estimation of vegetation fraction”, *International of Remote Sensing*, 23(13), 2537-2562.

Goldberg D., Holland J. H., (1988), “Genetic algorithms and machine learning”, *Machine Learning*, 3 (2), 95-99.

Gong P., Wang j., Yu L., Zhao Y., Zhao Y., Liang L., Niu Z., Huang X., Fu H., Liu S., Li C., Li X., Fu W., Liu C., Xu Y., Wang X., Cheng Q., Hu L., Yao W., Zhang H., Zhu P., Zhao Z., Zhang H., Zheng Y., Ji L., Zhang Y., Chen H., Yan A., Guo J., Yu L., Wang L., Liu X., Shi T., Zhu M., Chen Y., Yang G., Tang P., Xu B., Giri C., Clinton N., Zhu Z., Chen J., Chen J., (2013), “Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data”, *International Journal of Remote Sensing*, 34 (7), 2607-2654.

Goward S. N., Markham B., Dye D. G., Dualney W., Yang J., (1991), “Normalized difference vegetation index measurements from the advanced very high resolution radiometer”, *Remote Sensing of Environment*, 35 (2-3), 257-277.

Grekousis G., Mountrakis G., Kavouras M., (2015), “An overview of 21 global and 43 regional land-cover mapping products”, *International Journal of Remote Sensing*, 36 (21), 5309-5335.

Guyon I., Weston J., Barnhill S., Vapnik V., (2002), “Gene selection for cancer classification using support vector machines”, *Machine Learning*, 46 (1), 389-422.

Haralick R. M., Shanmugam K., Dinstein I., (1973), “Textural features for image classification”, *IEEE Transactions on Systems, Man, and Cybernetics*, 3 (6), 610-621.

Hatfield J. L., Kanemasu E. T., Asrar G., Jackson R. D., Pinter P. J., Reginato R. J., Idso S. B., (1985), “Leaf-area estimates from spectral measurements over various planting dates of wheat”, *International Journal of Remote Sensing*, 6 (1), 167-175.

He D., Rish I., Parida L., (2014), “Transductive HSIC Lasso”, 14th SIAM International Conference on Data Mining, 154-162, Philadelphia, Pennsylvania, USA, 24-26 April.

Hira Z. M., Gillies D. F., (2015), “A review of feature selection and feature extraction methods applied on microarray data”, *Advances in Bioinformatics*, 2015, 198363.

Huang C., Davis L. S., Townshend J. R. G., (2002), “An assessment of support vector machines for land cover classification”, *Internatioanl Journal of Remote Sensing*, 23 (4), 725-749.

Huete A., (1988), “A soil-adjusted vegetation index (SAVI)”, *Remote Sensing of Environment*, 25 (3), 295-309.

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.

Hughes G. F., (1968), "On the mean accuracy of statistical pattern recognizers", *IEEE Transactions on Information Theory*, 14 (1), 55-63.

Hunt E. R., Rock B. N., (1989), "Detection of changes in leaf water content using Near-and Middle-Infrared reflectances", *Remote Sensing of Environment*, 30 (1), 43-54.

Hurskainen P., Adhikari H., Siljander M., Pellikka P. K. E., Hemp A., (2019), "Auxiliary datasets improve accuracy of object based land use/land cover classification in heterogeneous savanna landscapes", *Remote Sensing of Environment*, 233, 111354.

Ibrahim, L., Samah A. A., Fauzi R., Noor N. M., (2016), "The land surface temperature impact to land cover types", *The International Archives of Photogrammetry, Remote Sensing and Spatial Informations Science*, 41, 871-876.

İşçi Ö., Korukoğlu S., (2003), "Genetik algoritma yaklaşımı ve yöneylem araştırmasında bir uygulama", *Yönetim ve Ekonomi Dergisi*, 10 (2), 191-208.

Jia S., Ji Z., Qian Y., Shen L., (2012), "Unsupervised band selection for hyperspectral imagery classification without manual band removal", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5 (2), 531-543.

Jin Y., Liu X., Chen Y., Liang X., (2018), "Land-cover mapping using random forest classification and incorporating NDVI time-series and texture: A case study of central Shandong", *International Journal of Remote Sensing*, 39 (23), 8703-8723.

Kavzoglu T., (2001), "An investigation of the design and use of feed-forward artificial neural networks in the classification of remotely sensed images", Doctor Thesis, The University of Nottingham.

Kavzoglu T., (2017), "Object-oriented random forest for high resolution land cover mapping using Quickbird-2 imagery". In: P. Samui, S. S. Roy, V. E. Balas, Editors, "Handbook of Neural Computation", Amsterdam: Elsevier, Chapter 33, 607-619.

Kavzoglu T., Bilucan F., Teke A., (2020), "Comparison of support vector machines, random forest and decision tree methods for classification of Sentinel-2A image using different band combinations", 41st Asian Conference on Remote Sensing, Deqing, China, 9-11 November.

Kavzoglu T., Colkesen I., (2009), "A kernel functions analysis for support vector machines for land cover classification", *International Journal of Applied Earth Observation and Geoinformation*, 11 (5), 352-359.

Kavzoglu T., Colkesen I., (2012), "The effects of training set size for performance of support vector machines and decision trees", 10th International Symposium on Spatial

Accuracy Assessment in Natural Resources and Environmental Sciences, 127-132, Catarina, Brazil, 10-13 July.

Kavzoglu T., Mather P. M., (2002), "The role of feature selection in artificial neural network application", *International Journal of Remote Sensing*, 23 (15), 2919-2937.

Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T. Y., (2017), "Lightgbm: A highly efficient gradient boosting decision tree", *Advances in Neural Information Processing Systems*, 3146-3154.

Khalid S., Khalil T., Nasreen S., (2014), "A survey of feature selection and feature extraction techniques in machine learning", *Science and Information Conference*, 372-378, London, UK, 27-29 August.

Khatami R., Mountrakis G., (2012), "Implications of classification of methodological decisions in flooding analysis from Hurricane Katrina", *Remote Sensing*, 4 (12), 3877-3891.

Kiala Z., Mutanga O., Odindi J., Peerbhay K., (2019), "Feature selection on Sentinel-2 multispectral imagery for mapping a landscape infested by parthenium weed", *Remote Sensing*, 11, 1892.

Kononenko I., (1994), "Estimating attributes: Analysis and extensions of RELIEF", *Proceedings of the 7th European Conference on Machine Learning*, 145-154, Catania, Italy, 5-8 April.

Ladha L., Deepa T., (2011), "Feature selection methods and algorithms", *International Journal on Computer Science and Engineering*, 3 (5), 1787-1797.

Landgrebe D., (2002), "Hyperspectral image data analysis", *IEEE Signal Processing Magazine*, 19 (1), 17-28.

Legleiter C. J., Marcus W. A., Lawrence R. L., (2002), "Effects of sensor resolution on mapping in-stream habitats", *Photogrammetric Engineering and Remote Sensing*, 68 (8), 801-807.

Lei S., (2012), "A feature selection method based on information gain and genetic algorithm", *International Conference on Computer Science and Electronics Engineering*, Hangzhou, Zhejiang, China, 355-358, 23-25 March.

Li J., Cheng K., Wang S., Morstatter F., Trevino R. P., (2017), "Feature selection: A data perspective", *ACM Computing Surveys*, 50 (6), 1-45.

Li S. J., Wu H., Wan D. S., Zhu J. L., (2011), "An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machines", *Knowledge-Based Systems*, 24 (1), 40-48.

Lillesand T. M., Kiefer R. W., Chipman J. W., (2015), "Remote sensing and image interpretation", 7th Edition, John Wiley & Sons.

Lloyd c. T., Sorichetta A., Tatem A. J., (2017), “High resolution global gridded data for use in population studies”, *Scientific Data*, 4 (1), 1-17.

Louhaichi M., Borman M. M., Johnson, D. E., (2001), “Spatially located platform and aerial photography for documentation of grazing impacts on wheat”, *Geocarto International*, 16 (1), 65-70.

Löw F., Michel U., Dech S., Conrad C., (2013), “Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines”, *ISPRS Journal of Photogrammetry and Remote Sensing*, 85, 102-119.

Maktav D., Erbek F. S., Jürgens C., (2005), “Remote sensing of urban areas”, *International Journal of Remote Sensing*, 26 (4), 655-659.

Marsett R. C., Qi J., Heilman P., Biedenbender S. H., Watson M. C., Amer S., Weltz M., Goodrich D., Marsett R., (2006), “Remote sensing for grassland management in the arid southwest”, *Rangeland Ecology & Management*, 59 (5), 530-540.

Mather P. M., Koch M., (2011), “Computer processing of remotely-sensed images: An introduction”, 4th Edition, Wiley-Blackwell.

Maxwell, A. E., Warner T. A., Fang F., (2018), “Implementation of machine learning classification in remote sensing: An applied review”, *International Journal of Remote Sensing*, 39 (9), 2784-2817.

Mingers J., (1989), “An empirical comparison of pruning methods for decision tree induction”, 4, 227-243.

Moghimi A., Yang C., Marchetto P. M., (2018), “Ensemble feature selection for plant phenotyping: A journey from hyperspectral to multispectral imaging”, *IEEE Access*, 6, 56870-56884.

Moore J. A., Primbblet, K. A., Drinkwater M. J., (2006), “Mathematical morphology: Star/galaxy differentiation & galaxy morphology classification”, *Publication of the Astronomical Society of Australia*, 23 (4), 135-146.

Mortoja M. G., Yigitcanlar T., (2020), “Local drivers of anthropogenic climate change: Quantifying the impact through a remote sensing approach in Brisbane”, *Remote Sensing*, 12 (14).

Mura M. D., Benediktsson J. A., Waske B., Bruzzone L., (2010), “Morphological attribute profiles for the analysis of very high resolution images”, *IEEE Transactions on Geoscience and Remote Sensing*, 48 (10), 3747-3762.

Mustapha M. R., Lim H. S., Mat M. Z., (2010), “Comparison of neural network and maximum likelihood approaches in image classification”, *Journal of Applied Sciences*, 10 (22), 2847-2854.

Myburgh G., van Niekerk A., (2014), "Impact of training set size on object-based land cover classification", *International Journal of Applied Geospatial Research*, 5 (39), 49-67.

Nagasubramanian K., Jones S., Sarkar S., Singh A. K., Singh A., Ganapathysubramanian B., (2018), "Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems", *Plant Methods*, 14 (1), 86.

Ozbakır A., Bannari A., (2008), "Performance of TDVI in urban land use/cover classification for quality of place measurement", *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, 691-694, Beijing, China, 3-11 July.

Pal M., (2005), "Random forest classifier for remote sensing classification", *International Journal of Remote Sensing*, 26 (1), 217-222.

Pal M., (2006), "Support vector machine-based feature selection for land cover classification: A case study with DAIS hyperspectral data", *International Journal of Remote Sensing*, 27 (14), 2877-2894.

Pal M., Mather P. M., (2005), "Support vector machines for classification in remote sensing", *International Journal of Remote Sensing*, 26 (5), 1007-1011.

Penuelas J., Baret F., Filella I., (1995), "Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectances", *Photosynthetica*, 31 (2), 221-230.

Pinty B., Verstraete M. M., "GEMI: A non-linear index to monitor global vegetation from satellites", *Vegetatio*, 101 (1), 15-20.

Plackett R. L., (1983), "Karl Pearson and the Chi-Squared Test", *International Statistical Review*, 51 (1), 59-72.

Polikar R., (2012), "Ensemble learning", In: C. Zhang, Y. Ma, Editors, "Ensemble machine learning", 1-34, Springer Science.

Pour A. B., Park Y., Park T. Y. S., Hong J. K., Hashim M., Woo J., Ayoobi I., (2018), "Regional geology mapping using satellite-based remote sensing approach in Northern Victoria Land, Antarctica", *Polar Science*, 16, 23-46.

Radhika K., Varadarajan S., (2016), "A tutorial on classification of remote sensing data", *International Research Journal of Engineering and Technology*, 3 (8), 881-885.

Rasul A., Balzter H., Ibrahim G., Hameed H., Wheeler J., Adamu B., Ibrahim S., Najmaddin P., (2018), "Applying built-up and bare-soil indices from Landsat 8 to cities in dry climates", *Land*, 7 (3), 81.

Richards J. A., Xiuping J., (1999), "Remote sensing digital image analysis", 4th Edition, Springer.

Rogan J., Miller J., Stow D., Franklin J., Levien L., Fischer C., (2003), "Land-cover change monitoring with classification trees using Landsat TM and ancillary data", *Photogrammetric Engineering and Remote Sensing*, 69 (7), 793-804.

Rondeaux G., Steven M., Baret F., (1996), "Optimization of soil-adjusted vegetation indices", *Remote Sensing of Environment*, 55 (2), 95-107.

Roujean J. L., Breon F. M., (1995), "Estimating PAR absorbed by vegetation from bidirectional reflectance measurements", *Remote Sensing of Environment*, 51 (3), 375-384.

Rwanga S. S., Ndambuki J. M., (2017), "Accuracy assessment of land use/land cover classification using remote sensing and GIS", *International Journal of Geosciences*, 8 (4), 611-622.

Saeyns Y., Inza I., Larranaga P., (2007), "A review of feature selection techniques in bioinformatics", *Bioinformatics*, 23 (19), 2507-2517.

Saha A. K., Arora M. K., Csaplovics E., Gupta R. P., (2005), "Land cover classification using IRS Liss III image and DEM in a rugged terrain: A case study in Himalayas", *Geocarto International*, 20 (2), 33-40.

Salehi B., (2012), "A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery" *International Journal of Geoinformatics*, 8 (2), 35-51.

Sertel E., Akay S. S., (2015), "High resolution mapping of urban areas using SPOT-5 images and ancillary data", *International Journal of Environment and Geoinformatics*, 2 (2), 63-76.

Shen L., Li C., (2010), "Water body extraction from Landsat ETM+ imagery using adaboost algorithm", *18th International Conference on Geoinformatics*, Kyiv, Ukraine, 13-16 May

Shippert P., (2004), "Why use hyperspectral imagery?", *Photogrammetric Engineering and Remote Sensing*, 70 (4), 377-396.

Singh D. A., Balamurugan S., Leavline E., (2016), "Literature review on feature selection methods for high-dimensional data", *International Journal of Computer Applications*, 136, 9-17.

Sisodia P. S., Tiwari V., Kumar A., (2014), "A comparative analysis of remote sensing image classification techniques", *Proceedings of the 2014 International Conference on Advanced in Computing, Communications and Informatics*, 1418-1421, New Delhi, India, 24-27 September.

Sluiter R., Pebesma E. J., (2010), "Comparing techniques for vegetation classification using multi- and hyperspectral images and ancillary environmental data", *International Journal of Remote Sensing*, 31 (23), 6143-6161.

Takahashi Y., Ueki M., Yamada M., Tamiya G., Motoike I. N., Saigusa D., Sakuraji

M., Nagami F., Ogishima S., Koshiha S., Kinoshita K., Yamamoto M., Tomita H., (2020), "Improved metabolomic data-based prediction of depressive symptoms using nonlinear machine learning with feature selection", *Translational Psychiatry*, 10 (1), 1-12.

Tang C., Luktarhan N., Zhao Y., (2020), "An efficient intrusion detection method based on lightgbm and autoencoder", *Symmetry*, 12 (9), 1-16.

Tavus B., Karataş K., Türker M., (2019), "Object-based crop pattern detection from IKONOS satellite images in agricultural areas", *Pamukkale University Journal of Engineering Sciences*, 25 (5), 603–614.

Thakkar A. K., Desavi V. R., Patel A., Potdar M.B., (2017), "Post-classification corrections in improving the classification of land use/land cover of arid region using RS and GIS: The case of Arjuni watershed, Gujarat, India", *Egyptian Journal of Remote Sensing and Space Science*, 20 (1), 79-89.

Thomas S., Kuska M. T., Bohnenkamp D., Brugger A., Alisaac E., Wahabzada M., Behmann J., Mahlein A. K., (2018), "Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective", *Journal of Plant Disease and Protection*, 125 (1), 5-20.

Topp S. N., Pavelsky T. M., Jensen D., Simard M., Ross M. R. V., (2020), "Research trends in the use of remote sensing for inland water quality science: Moving towards multidisciplinary applications", *Water*, 12 (1), 169.

Tso B., Mather P. M., (2009), "Classification methods for remotely sensed data", Second Edition, Boca Raton, CRC Press.

Tsoeleng L., Odindi J., Mhangara P., (2020), "A comparison of two morphological techniques in the classification of urban land cover", *Remote Sensing*, 12 (7), 1089.

Tuia D., Pacifici F., Kanevski M., Emery W. J., (2009), "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines", *IEEE Transactions on Geoscience and Remote Sensing*, 47 (11), 3866-3879.

Ustuner M., Sanli F. B., (2019), "Polarimetric target decompositions and light gradient boosting machine for crop classification: A comparative evaluation", *ISPRS International Journal of Geo-Information*, 8 (2), 97.

Üstüner M., Abdikan S., Bilgin G., Şanlı F. B., (2020), "Hafif gradyan artırma makineleri ile tarımsal ürünlerin sınıflandırılması", *Türk Uzaktan Algılama ve CBS Dergisi*, 1 (2), 97-105.

Van Deventer A. P., Ward A. D., Gowda P. M., Lyon J. G., (1997), "Using thematic mapper data to identify contrasting soil plains and tillage practices", *Photogrammetric Engineering and Remote Sensing*, 63 (1), 87-93.

Varma M. K. S., Rao N. K. K., Raju K. K., Varma G. P. S., (2016), “Pixel-based classification using support vector machine classifier”, 6th International Conference on Advanced Computing, 51-55, Bhimavaram, India, 27-28 February.

Wang J., Huang B., Zhang H. K., Ma P., (2019), “Sentinel-2A image fusion using a machine learning approach”, IEEE Transactions on Geoscience and Remote Sensing, 57 (12), 9589-9601.

Waqar M. M., Mirza J. F., Mumtaz R., Hussain E., (2012), “Development of new indices for extraction of built-up area & bare soil from Landsat data”, Open Access Scientific Reports, 1 (1), 4.

Web 1, (2021), https://cdnacikogretim.istanbul.edu.tr/auzefcontent/20_21_Guz/uz-aktan_algilama/2/index.html#konu-2, (Erişim tarihi: 01/05/2021).

Web 2, (2021), https://tr.wikipedia.org/wiki/Dosya:Eletromanyetik_dalga_tayf%C4%B1.png, (Erişim tarihi: 05/05/2021).

Web 3, (2021), <https://hoasys.fr/en/products/modular-hyper-spectral-camera>, (Erişim tarihi: 10/05/2021).

Web 4, (2021), <https://www.worldpop.org/>, (Erişim 13/04/2021).

Whitney, A. W., (1971), “A direct method of nonparametric measurement selection”, IEEE Transactions On Computers, 20, 1100-1103.

Yamada M., Jitkrittum W., Sigal L., Xing E. P., Sugiyama M., (2014), “High-dimensional feature selection by feature-wise kernelized Lasso”, Neural Computation, 26 (1), 185-207.

Yang H., (1999), “A back-propagation neural network for mineralogical mapping from AVIRIS data”, International Journal of Remote Sensing, 20(1), 97-110.

Zeferino L. B., de Souza L. F. T., do Amaral C. H., Filho E. I. F., de Oliveira T. S., (2020), “Does environmental data increase the accuracy of land use and land cover classification?”, International Journal of Applied Earth Observation and Geoinformation, 91, 102128.

Zhang D., Gong Y., (2020), “The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure”, IEEE Access, 8, 220990-221003.

Zhang R., Ma J., (2009), “Feature selection for hyperspectral data based on recursive support vector machines”, International Journal of Remote Sensing, 30(14), 3669-3677.

Zhou Y., Zhang R., Wang S., Wang F., (2018), “Feature selection method based on high-resolution remote sensing images and the effect of sensitive features on classification accuracy”, Sensors, 18(7).

Zhu Z., Gallant A. L., Woodcock C. E., Pengra B., Olofsson P., Loveland T. R., Jin S., Dahal D., Yang L., Auch R. F., (2016), "Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative", ISPRS Journal of Photogrammetry and Remote Sensing, 122, 206-221.

