

# NESNE TABANLI GÖRÜNTÜ ANALİZİNDE ÖLÇEK VE EĞİTİM SETİ BOYUTUNUN SINIFLANDIRMA DOĞRULUĞUNA ETKİLERİNİN ARAŞTIRILMASI

İ. Aşık<sup>a</sup>, İ. Çölkesen<sup>b</sup>

<sup>a</sup>Ardahan Çevre ve Şehircilik İl Müdürlüğü 75000 Merkez-Ardahan(ismail.asik@csb.gov.tr)

<sup>b</sup>Gebze Teknik Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, 41400 Gebze-Kocaeli (icolkesen@gtu.edu.tr)

**ANAHTAR SÖZCÜKLER:** Objeye Tabanlı Sınıflandırma, Segmentasyon, Ölçek Parametresi, ESP-2, Moran's-I.

## ÖZET:

Son yıllarda, yüksek konumsal çözünürlüğe sahip uydu görüntülerinin sınıflandırılmasında klasik piksel tabanlı sınıflandırma yaklaşımı yerine nesne tabanlı sınıflandırma yaklaşımı tercih edilmeye başlanmıştır. Nesne tabanlı sınıflandırma analizlerinde benzer piksellerin bir grup altında bir araya getirilmesi için kullanılan segmentasyon işleminin gerçekleştirilmesinde, üretilecek görüntü nesnesi boyutunu belirleyen ölçek parametresinin tespiti önemli bir husustur. Bununla birlikte nesne tabanlı sınıflandırma bir kontrollü sınıflandırma tekniği olduğundan kullanılacak algoritmanın eğitimi ve testi için kullanılacak örnek sayısı sınıflandırma doğruluğuna etki edebilecek diğer önemli bir husustur. Bu çalışmada, segmentasyon parametreleri ve eğitim veri seti boyutu arasındaki ilişki incelenmiştir. Temel veri seti olarak 2018 tarihli Sentinel-2 uydu görüntüsü kullanılmıştır. Görüntü segmentasyonu için ölçek parametrelerini belirlemede ESP-2 ve Moran's I yöntemi kullanılmıştır. ESP-2 algoritması ile muhtemel ölçek parametreleri 32, 59 ve Moran's I yöntemiyle de 20 olarak belirlenmiştir. Örneklem oranının nesne tabanlı sınıflandırma doğruluğu üzerindeki etkisini değerlendirmek ve her ölçek parametresi için en iyi örneklem oranını belirlemek için, sekiz örneklem oranı (10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20) göz önünde bulundurularak eğitim ve test veri setleri oluşturulmuştur. Oluşturulan segmentlerin sınıflandırılmasında ve tematik harita üretiminde rastgele orman, en yakın komşuluk, karar ağaçları, AdaBoost algoritmalarından yararlanılmıştır. Sonuçlar, AdaBoost'un algoritmasının sınıflandırma performansının, her bir ölçek parametresi ve sınırlı eğitim örnekleri için NN, DT ve RF performanslarına göre daha üstün olduğunu gösterdi. En yüksek toplam doğruluk Adaboost ile ölçek parametresi 59 ve örneklem oranı 30:70 için %95,30 olarak hesaplanmıştır. Eğitim veri seti boyutundaki artış ile sınıflandırma performansının belirli bir düzeye kadar arttığı gözlemlenmiştir. Çalışma sonuçları ölçek parametresi ve eğitim veri seti boyutundaki değişimlerin objeye tabanlı sınıflandırma doğruluğuna etkileri açısından önemli bilgiler sağlayacaktır.

**KEY WORDS:** Object-Based Classification, Segmentation, Scale Parameters, ESP-2, Moran's-I.

## ABSTRACT:

In recent years, instead of traditional pixel-based classification approach, object-based classification approach is preferred for especially classification of the high spatial resolution satellite imagery. Determination of the optimal scale parameter identifying image object size for segmentation process aiming to combine pixels having similar characteristics into a group is an important issue in performing effective object-based image analysis. Due to object-based method is a supervised image classification technique, the number of samples to be used for the training of the algorithm is another important issue that can affect the classification accuracy. In this study, the relationship between segmentation parameters (i.e. scale parameter) and the size of the training dataset was investigated. The Sentinel-2 satellite image acquired in 2018 was used as a fundamental input dataset. The ESP-2 and Moran's I methods were used to determine the optimal scale parameters for image segmentation process. The optimal scale parameters were estimated by the ESP-2 algorithm as 32 and 59, while the scale parameter was determined as 20 with the Moran's I method. In order to assess the effect of sampling ratio on object-based classification accuracy and to determine the best sampling ratio for each scale parameter, training and testing datasets were formed considering the eight sampling ratios as 10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30 and 80:20. For the classification of segmented image objects and to produce land use and land cover thematic map, nearest neighbour (NN), decision trees (DT), AdaBoost and random forest (RF) algorithms were used. Results showed that the classification performance of the Adaboost algorithm was found to be superior compared to those of NN, DT and RF for each scale parameter and for the case of limited training samples. The highest overall accuracy was estimated with Adaboost as 95.30% for scale parameter 59 and sampling ratio of 30:70. Moreover, it was observed that the performance of the algorithms increased to a certain level with the increase in the size of the training dataset. The results of the study will provide important information on the effects of scale parameters and changes in the size of the training dataset to the object-based classification accuracy.

<sup>a</sup> ismail.asik@csb.gov.tr

<sup>b</sup> icolkesen@gtu.edu.tr

## 1. GİRİŞ

Son yıllarda artan uydu teknolojilerindeki gelişmeler beraberinde uzaktan algılama teknolojileri ve uydu görüntülerinin sahip olduğu özelliklerin de iyileşmesine neden olmuştur. Uydu görüntüleri aracılığıyla arazi kullanımını temsil eden tematik haritalar hızlı, ucuz ve doğru bir şekilde üretilmekte ve analiz edilebilmektedir (Bisht and Kathyari, 2001; Dedeoğlu ve diğ., 2016). Uzaktan algılamada görüntü sınıflandırması, uydu görüntüsündeki her bir pikselin yeryüzünde karşılık geldiği arazi örtüsü sınıfına atanması işlemi olarak tanımlanabilir. Uydu görüntüsünün sınıflandırılması neticesinde farklı ölçeklerde yeryüzünü temsil eden arazi örtüsü ve arazi kullanım haritaları üretilmekte ve söz konusu haritalar yeryüzü ile ilişkili birçok çalışmada temel veri kaynağı olarak kullanılmaktadır. Kullanılan uydu görüntüsünün özelliklerine göre sınıflandırma işlemi için seçilen yöntem ve uygulanan algoritmalar doğruluğu etkileyen faktörlerdir.

Uydu görüntülerinin sınıflandırılmasında genel olarak piksel tabanlı ve nesne tabanlı sınıflandırma olmak üzere iki yaklaşım bulunmaktadır. Piksel tabanlı sınıflandırma, uydu görüntüsünün temel birimi olan piksellerin parlaklık değerleri esas alınarak ayrı ayrı sınıflandırılması söz konusudur. Özellikle yüksek mekânsal çözünürlüklü uydu görüntülerinin kullanımıyla birlikte görüntü veri seti boyutu artmış ve benzer spektral özelliklere sahip piksellerin birbirinden ayırt edilerek sınıflandırılması güçleşmiştir. Bu zorlukların üstesinden gelebilmek ve yüksek çözünürlüklü görüntüler yardımıyla doğruluğu yüksek tematik harita üretiminde klasik piksel tabanlı sınıflandırma yerine nesne tabanlı sınıflandırma yaklaşımının kullanımı yaygın hale gelmiştir. Nesne tabanlı sınıflandırma, uydu görüntüsündeki pikselleri değil benzer spektral özelliklere sahip piksellerin bir araya gelmesiyle oluşan nesnelere (objelerin) sınıflandırılması işlemi yapmaktadır.

Nesne tabanlı sınıflandırmada, görüntü segmentasyonu, görüntü nesnelere özelliklerinin belirlenmesi ve görüntü nesnelere sınıflandırılması olmak üzere üç temel işlem adımı bulunmaktadır. Bunlardan ilki ve nesne tabanlı sınıflandırmanın temel işlem adımı olan ve benzer spektral özelliklere sahip piksellerin bir araya getirilerek görüntü nesnelere oluşturulması olarak bilinen segmentasyon işlemidir. Ölçek parametresi, segmentasyon ile elde edilen nesnelere boyutunun belirlenmesinde en önemli faktördür (Bo and Han, 2010; Li and Shao, 2012; Kavzoğlu ve Yıldız, 2015). Bu amaçla en uygun segmentasyon parametresinin tespiti nesne tabanlı sınıflandırma işleminin doğruluğunu doğrudan etkileyen önemli bir husustur (Kim et al., 2011; Myint et al., 2011; Dragut et al., 2014; Ma et al., 2015). Ölçek parametresi belirlemede farklı yaklaşımlar kullanılmaktadır. Bunlardan en yaygın kullanılanı, kullanıcı tarafından anlamlı görüntü nesnelere elde edilene kadar ölçek parametresinin manuel olarak ayarlanmasıdır (deneme

yanılma yolu). Açıkçası bu yaklaşım kullanıcının tecrübe ve deneyimine bağlı olduğundan tercih edilmemektedir. Literatürde ölçek parametresini otomatik veya yarı otomatik tespiti için farklı yöntemler önerilmiştir (Ma et al., 2015). Drağut et al. (2014) yılında en uygun ölçek parametresini belirlemek için ESP2 yöntemini geliştirmişlerdir. Ölçek parametresini belirlemede kullanılan diğer bir yöntem konumsal otokorelasyon tekniği olan Moran's İndeksi ve varyans değerinin birlikte kullanılmasıdır (Kavzoğlu ve Yıldız, 2015).

Nesne tabanlı sınıflandırma bir kontrollü sınıflandırma tekniği olduğundan, sınıflandırma doğruluğuna etki eden diğer bir etken ise sınıflandırma modelinin oluşturulmasında kullanılacak eğitim veri seti boyutudur (Pal and Mather, 2003; Foody et al., 2006). Eğitim veri seti boyutu belirlemede kullanılan standart bir yaklaşım bulunmamakta olup güncel araştırma konusudur.

Bu çalışmada, Sentinel-2a uydu görüntüsü kullanarak nesne tabanlı sınıflandırma yaklaşımıyla çalışma alanı olarak belirlenen Sakarya ili Ferizli ilçesine ait arazi kullanımını ve arazi örtüsünü temsil eden tematik haritanın üretilmesi problemi ele alınmıştır. Çalışmanın genel amacı, segmentasyon işlemi için seçilecek farklı ölçek parametreleri ve eğitim veri seti boyutundaki değişimlerin sınıflandırma doğruluğuna etkilerinin araştırılmasıdır. Bu amaçla ölçek parametresi tespitinde Moran's İndeksi ve ESP yöntemleri kullanılırken, her bir ölçek parametresi için 8 farklı boyutta eğitim veri seti kullanılmıştır. Algoritmalar ile her ölçek ve eğitim veri seti için doğruluk analizi yapılmış ve sonuçları karşılaştırılmıştır.

## 2. ÇALIŞMA ALANI VE KULLANILAN VERİ

Bu çalışmada Sakarya İlinin kuzeyinde kalan Ferizli ve Karasu İlçeleri ile Sakarya Nehrinin çevrelediği Limandere, Akgöl, Sinanoğlu ve Adatepe mahalle ve köylerini kapsayan bölge çalışma alanı olarak belirlenmiştir. Bölgede yoğunlukla fındık, mısır başta olmak üzere diğer tarım ürünleri yetiştiriciliği yapılmaktadır. Çalışmada bölgenin arazi örtüsü ve arazi kullanımını belirlenmesi amacıyla 2018 yılında kaydedilen Sentinel-2 uydu görüntüsü kullanılmıştır. Sentinel-2 uydusu 13 banttan oluşmaktadır. Bu çalışmada Sentinel-2 uydu görüntüsünün 10m çözünürlüğe sahip 4 bantı (bant-2, bant-3, bant-4 ve bant-8) ve 20m çözünürlüğündeki altı bantı (bant-5, bant-6, bant-7, bant-8A, bant-11 ve bant-12) kullanılmıştır. 20m konumsal çözünürlüğe sahip bantlar seçilen 10m çözünürlüklü bantlar kullanılarak Gram-Schmidt algoritması yardımıyla pan-sharpen işlemi yapılmıştır. Bunun gerçekleştirilmesinde Selva et al. (2005) tarafından önerilen yaklaşım kullanılmıştır. Bu yaklaşım, pan-sharpen işleminde kullanacak 10m çözünürlüklü bant seçiminde 20m bant ile en yüksek korelasyona sahip olanın belirlenmesi esasına dayanmaktadır. Çalışma kapsamında 20m çözünürlüğündeki bant-6, 7 ve 8A bantları için 10m çözünürlüğündeki bant-4 pan-sharpen işleminde kullanılırken; 20m çözünürlüğündeki 5, 11 ve 12 bantları için 10m çözünürlüğündeki bant-8 kullanılmıştır. Çalışma alanı ve Sentinel-2 uydu görüntüsü Şekil 1'de görülmektedir.



Şekil 1. Çalışma alanı ve uydu görüntüsü.

### 3. OBJE TABANLI SINIFLANDIRMA

Obje tabanlı sınıflandırma 3 aşamadan oluşmaktadır. Birinci aşama uydu görüntüsünün segmentasyonu, ikinci aşama veri setinin oluşturulması ve son aşama sınıflandırma ve analiz olarak ifade edilebilir.

Obje tabanlı sınıflandırma, piksel tabanlı sınıflandırmanın aksine tek piksel yerine piksellerin bir araya gelmesiyle oluşan objeleri dikkate alacak şekilde geliştirilmiştir. Görüntü segmentasyonu için eCognition yazılımında birçok teknik geliştirilmiştir. Bunlar; satranç tahtası, dörtlü ağaç, spektral farklılık, kontrast ayırma, çok eşik segmentasyonu ve çoklu çözünürlüklü segmentasyon (MRS) teknikleridir (Kavzoğlu ve diğ., 2015). Literatür incelendiğinde sıklıkla kullanılan segmentasyon tekniği olarak çoklu çözünürlüklü segmentasyon yöntemi bu çalışmada kullanılmıştır. Çoklu çözünürlüklü segmentasyon işleminde objelerin oluşmasında kullanıcı tarafından belirlenmesi gereken parametreler; ölçek, bütünlük ve şekil parametreleridir. Literatürde, söz konusu parametreler arasında nesne tabanlı görüntü analizi için en önemlisinin ölçek parametresi olduğu ifade edilmektedir (Kavzoğlu ve diğ., 2015; Myint et al., 2011; Belgiu et al., 2014). Bu çalışma kapsamında ölçek parametresinin tespitinde ESP-2 (Estimation of Scale Parameter) ve Moran's-I yöntemi kullanılmıştır.

Uydu görüntüsünün çok bantlı olması ve tüm bantları işleme almasından dolayı ESP aracı yerine Dragut et al. (2014) tarafından geliştirilen ve lokal varyansın istatistiksel analizine dayanan ESP-2 aracı kullanılmıştır. ESP-2 aracıyla ölçek parametresi tespitinde, ölçek parametresi değişimine göre hesaplanan lokal varyanslar ve değişimi gösteren lokal varyans grafiği kullanılmaktadır. Lokal varyans grafikleri, uydu görüntüsünün veri özelliklerine uygun ölçek parametrelerini değerlendirmektedir (Tonbul ve Kavzoğlu, 2018). Ölçek parametrelerine karşılık gelen lokal varyans değerleri ile değişim oranı hesaplanıp LV-

ROC grafiği elde edilmektedir. Bu grafikteki ani değişimlerin olduğu pikler uydu görüntüsünün optimum ölçek parametre değerini belirlemektedir (Dragut et al., 2010). Uygulama kapsamında ölçek parametresi tespitinde kullanılan diğer bir yöntem ise Moran's-I yöntemidir. Konumsal otokorelasyon tekniğine olarak bilinen Moran's-I konumsal objeler arasındaki istatistiksel ayrımın gerçekleştirilmesinde kullanılan bir yöntemdir (Fatheringham et al., 2000; Kavzoğlu ve Yıldız, 2015). Moran's I birçok uygulamada segmentasyon niteliğinin tayininde değerlendirilmiştir (Espindola et al., 2006; Kim et al., 2008).

## 4. SINIFLANDIRMA ALGORİTMALARI

### 4.1 Karar Ağaçları Algoritması

Karar ağaçları, sınıflandırma işlemindeki başarısı ve kullanılan karar kurallarının kullanıcılar tarafından açık bir şekilde anlaşılmasından dolayı son yıllarda en sık kullanılan sınıflandırma algoritmalarındandır. Karar ağaçlarının temel çalışma prensibi, kompleks yapıdaki bir sınıflandırma işleminin çözülmesinde problemi çok aşamalı bir hale getirerek işlemi gerçekleştirmesidir (Safavian and Langrebe, 1991).

Karar ağacı kök düğüm, dal ve yaprak olmak üzere 3 temel bölümden oluşur (Quinlan, 1993). Her uydu görüntüsü için bant değeri bir düğüm tarafından simgelenir (Kavzoğlu ve Çölkesen, 2010). Karar ağacının oluşumunda ana ilke, verilere bağlı bir dizi soruların sorulması ve elde edilen cevaplara göre karar kuralları oluşturmasıdır. Karar ağacının kök düğümünde verilerin sınıflandırılması ve ağaç yapısının oluşturulması için sorular sorulmaya başlanır ve dalları olmayan düğümler ya da yapraklar tespit edilene kadar işlem devam eder (Pal and Mather, 2003).

### 4.2 En Yakın Komşuluk Algoritması

Parametrik bir algoritma olan en yakın komşuluk (EYK), sınıflandırmada kullanılan yaygın bir metottur. EYK'nın çalışma prensibi sınıf etiketi belli olan piksellerden veya görüntü objelerden faydalanarak sınıfı belli olmayan piksellerin ve objelerin hangi sınıfa ait olduğunun belirlenmesinde bir uzaklık fonksiyonu (örneğin öklit uzaklığı) kullanarak spektral uzaklıkları kullanılmaktadır (Lillesand et al., 2007).

### 4.3 Rastgele Orman Algoritması

Makine öğrenme algoritması olan rastgele ormanın (RO) temeli karar ağaçlarına dayanmaktadır (Breiman, 2001; Kavzoğlu et al., 2014). RO, eğitim verilerini kullandığından kontrollü sınıflandırma yöntemi olarak adlandırılmaktadır. RO'nun çalışma prensibi karar ağacını oluşturmak için belirlediğimiz eğitim veri setinden rastgele seçim yaparak alt kümeleri oluşturur. Oluşturulan alt kümelerin 2/3'ü karar ağacını oluşturmada geri kalan 1/3 kısmında oluşturulan karar ağaç yapısını test etmek için kullanılır.

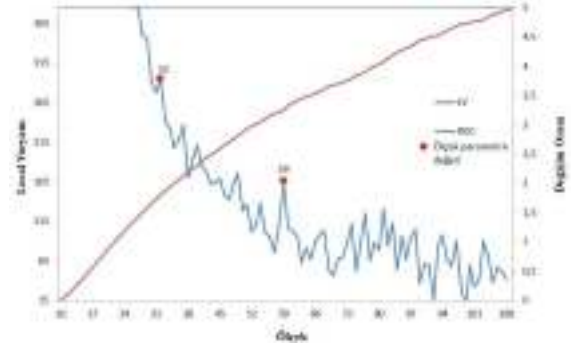
#### 4.4 AdaBoost Algoritması

AdaBoost algoritması birçok sınıflandırma probleminde kullanılan bir topluluk öğrenme algoritmasıdır. Algoritmanın temel çalışma prensibi iteratif olarak temel sınıflandırıcı olarak seçilen algoritmanın performansına bağlı olarak eğitim veri setinin yeniden düzenlenmesidir. Başlangıç olarak eğitim veri setindeki her bir örnek eşit ağırlıklı kabul edilerek, iteratif olarak sınıflandırma işlemi yapılır. Her bir sınıflandırma sonucunda hatalı sınıflandırılan örnekler eğitim veri setinde tekrar edilerek (ağırlığı artırılarak) sınıflandırma işlemi belirli bir seviyeye ulaşıncaya kadar devam eder. Hatalı sınıflandırılan örnekler yoğunlaşarak sınıflandırma doğruluğunun artırılmasını hedeflemektedir (Çölkesen ve Kavzoğlu, 2017).

#### 5. UYGULAMA

Bu çalışmada nesne tabanlı sınıflandırma yaklaşımıyla Sentinel-2 uydu görüntüsünün sınıflandırılması ve çalışma alanı olarak belirlenen tarımsal faaliyetlerin gerçekleştirildiği Ferizli ilçesine ait arazi örtüsü/arazi kullanımını temsil eden tematik harita üretilmesi amaçlanmıştır. Bu amaç doğrultusunda, çalışma alanını temsil eden arazi örtüsü ve arazi kullanım sınıfları tespit edilmiştir. Neticede, çalışma alanında mısır, su, orman, yol, fındık, yerleşim yeri, toprak ve ekili olmayan arazi olmak üzere 8 arazi örtüsü/kullanım sınıfı belirlenmiştir. Sınıfların belirlenmesinde uydu görüntüleri ve tematik haritalardan yararlanılmıştır.

Nesne tabanlı sınıflandırma işleminde eCognition yazılımından yararlanılmıştır. Segmentasyon algoritması olarak çoklu çözünürlüklü segmentasyon kullanılmıştır. Bu çalışmada optimum ölçek parametresi belirlemede ESP-2 ve Moran's *I* araçlarından yararlanılmıştır. ESP-2 yardımıyla optimum ölçek parametresi belirlemede ölçek aralığı olarak 10 ile 110 arasında 1 artışla ölçek değerleri için varyans değerlerindeki değişim analiz edilmiştir. Belirlenen ölçek parametrelerine karşılık gelen lokal varyans değerinin değişim oranı Şekil 2'de gösterilmektedir. Değişim oranı grafiğine göre optimum ölçek parametreleri belirlemek için grafikteki ani pik değerleri esas alınmıştır. Şekil 2'deki grafiğe bakılarak optimum ölçek parametre değerleri olarak 32 ve 59 değeri belirlenmiştir.



Şekil 2. ESP2 aracı ile ölçek parametre grafiği

Moran's *I* yönteminde, ölçek değeri 5'er artışla 5 ile 100 arasında toplam 20 ölçek değeri belirlenmiş ve her bir ölçek değeri esas alınarak segmentasyon işlemi yapılmıştır. Her bir segmentasyon sonucu oluşan görüntü objeleri için Moran's *I* ve varyans hesaplanmış olup, hesaplamalarda uydu görüntüsünün NIR (bant 7), kırmızı (bant 3) ve yeşil (bant 2) bantlar değerlendirilmiştir. Moran's *I* ve varyans değerleri farklı değerlere sahip olduklarından normalizasyon işlemi gerçekleştirilerek aynı değer aralığına çekilmiştir. Kullanılan her bant için normalize edilen değerler toplamı yani normalize varyans ile normalize Moran's *I* değerinin toplamı genel skor değerini vermiştir. 3 bant için belirlenen genel skor değerinin toplamının ortalaması oluşturulan segmentin genel skor değerini vermektedir. Genel skor değerinin minimum olduğu değer optimum ölçek parametre değerine denk gelmektedir (Tablo1) (Espindola et al., 2006; Kavzoğlu ve Yıldız, 2015).

Tablo 1. Moran's *I* yöntemiyle optimum ölçek parametre değeri

Ölçek	Kullanılan Bantlar			Toplam GS/3
	NIR GS	Kırmızı GS	Yeşil GS	
5	1,00	1,00	1,00	1
10	0,88	0,87	0,87	0,875197
15	0,87	0,81	0,80	0,827879
20	0,89	0,79	0,77	0,817224
25	0,91	0,78	0,76	0,817878
30	0,92	0,79	0,77	0,826404
35	0,93	0,81	0,79	0,839941
40	0,93	0,82	0,80	0,850599
45	0,94	0,83	0,80	0,856437
50	0,95	0,81	0,79	0,85378
55	0,95	0,83	0,82	0,86482
60	0,95	0,83	0,81	0,867296
65	0,96	0,84	0,83	0,878039
70	0,94	0,87	0,85	0,888909
75	0,96	0,88	0,87	0,900871
80	0,96	0,89	0,88	0,910151
85	0,97	0,91	0,90	0,92914
90	0,98	0,95	0,94	0,954719
95	0,99	0,97	0,97	0,978511
100	1,00	1,00	1,00	1

Her iki yöntemle belirlenen ölçek parametreleri, şekil ve yoğunluk parametreleri ve segmentasyon sonucu üretilen görüntü nesnelere ait sayıları Tablo 2’de verilmiştir. Belirlenen optimum ölçek parametre değerleri ve çoklu çözünürlüklü segmentasyon algoritmasıyla segmentler oluşturulmuştur. Oluşan nesnelere ait birçok temel özellik tanımlanabilmektedir. Bu çalışmada spektral özellikler olan minimum piksel, maksimum piksel, standart sapma, ortalama, oranlama ve NDVI hesaplanmıştır. Alan bazlı çalışma olduğundan dolayı geometrik özellik olan alan özelliği de hesaplanmıştır.

Tablo 2. Uygulamada kullanılan parametreler

	Ölçek	Şekil	Yoğunluk	Toplam Obje
Morans-I	20	0.1	0.5	27703
ESP	32	0.1	0.5	10762
	59	0.1	0.5	3235

Sınıflandırma işleminde kullanılmak üzere segmente edilen görüntü üzerinden eğitim ve test veri seti alanları seçilmiştir. Bu çalışma kapsamında veri seti boyutundaki değişimin sınıflandırma doğruluğuna etkisi araştırıldığından belirlenen oranlarda (10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20) eğitim ve test veri setleri oluşturulmuştur. Tablo 3’te eğitim ve test veri değerleri gösterilmiştir.

Tablo 3. Uygulamada kullanılan ölçek parametreleri ve eğitim veri seti boyutları

Ölçek	Eğitim veri seti boyutu oranları (%)									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	99%
Ölçek-10	Özellik	30	22	23	22	27	28	24	22	22
	Alan	30	29	27	27	27	27	27	27	27
	Yal	10	11	12	12	12	12	12	12	12
	Şekil	10	11	12	12	12	12	12	12	12
	Yoğunluk	10	11	12	12	12	12	12	12	12
	Toplam	10	11	12	12	12	12	12	12	12
	Özellik	11	11	12	12	12	12	12	12	12
	Alan	11	11	12	12	12	12	12	12	12
	Yal	11	11	12	12	12	12	12	12	12
	Şekil	11	11	12	12	12	12	12	12	12
Ölçek-20	Özellik	12	12	12	12	12	12	12	12	12
	Alan	12	12	12	12	12	12	12	12	12
	Yal	12	12	12	12	12	12	12	12	12
	Şekil	12	12	12	12	12	12	12	12	12
	Yoğunluk	12	12	12	12	12	12	12	12	12
	Toplam	12	12	12	12	12	12	12	12	12
	Özellik	12	12	12	12	12	12	12	12	12
	Alan	12	12	12	12	12	12	12	12	12
	Yal	12	12	12	12	12	12	12	12	12
	Şekil	12	12	12	12	12	12	12	12	12
Ölçek-30	Özellik	12	12	12	12	12	12	12	12	12
	Alan	12	12	12	12	12	12	12	12	12
	Yal	12	12	12	12	12	12	12	12	12
	Şekil	12	12	12	12	12	12	12	12	12
	Yoğunluk	12	12	12	12	12	12	12	12	12
	Toplam	12	12	12	12	12	12	12	12	12
	Özellik	12	12	12	12	12	12	12	12	12
	Alan	12	12	12	12	12	12	12	12	12
	Yal	12	12	12	12	12	12	12	12	12
	Şekil	12	12	12	12	12	12	12	12	12

Örnek Tablo 3 incelendiğinde seçilen farklı boyutlara göre toplanan eğitim veri seti değerleri analiz edildiğinde ölçek değeri büyüdükçe azalan segmentasyon sayısına bağlı seçilen eğitim veri seti boyutlarının da azaldığı görülmektedir. Ölçek değerinin 20 olduğunda tüm sınıflar için 2578 eğitim ve test veri seti örneği toplanmıştır. 32 ölçek değerinde segmentasyon sayısına bağlı olarak 981 örnek veri seti toplanmıştır. Son ölçek değeri olan 59 ölçeğinde ise 481 eğitim ve test veri seti örneği toplanmıştır.

Sınıflandırma algoritmaları olarak en yakın komşuluk (EYK), rastgele orman (RO), karar ağacı (KA) ve AdaBoost kullanılarak nesne tabanlı sınıflandırma

yapılmıştır. En yakın komşuluk algoritmasının genel doğruluk sonuçları örnek Tablo 6’da her ölçek için ayrı ayrı analiz verilmiştir. Ölçek parametresinin 20 olarak seçildiği durumda elde edilen sınıflandırma sonuçları incelendiğinde %10 eğitim veri setinden %80 eğitim veri seti oranına kadar genel doğruluğun yaklaşık %5 değerinde arttığı görülmüştür. 32 ve 59 ölçek değerleri incelendiğinde %10’luk eğitim veri setinde yaklaşık aynı genel doğruluğu verdiği gözlenmiştir. Eğitim veri setinde ki artışla doğruluğun %1-2 değerlerinde arttığı görülmüştür. EYK algoritmasının sonuçları bir bütün olarak analiz edildiğinde en iyi sonucu 20 ölçek değerinde %80 eğitim veri seti boyutunda verdiği sonucuna ulaşmıştır. Elde edilen bu sonuç sınırlı sayıda eğitim örneği için EYK algoritmasının performansının düşük olduğunu, eğitim veri setinin boyutunun artışına paralel olarak EYK algoritmasının da arttığı görülmektedir.

Tablo 6. EYK algoritmasının genel doğruluk değerleri

Eğitim:Test Oranı (%)	Ölçek-20	Ölçek-32	Ölçek-59
10:90	88.90	83.71	83.07
20:80	88.85	84.99	85.36
30:70	89.83	86.02	91.17
40:60	91.05	87.06	88.76
50:50	92.08	87.79	92.06
60:40	91.27	86.92	87.12
70:30	92.61	86.85	91.93
80:20	93.27	81.49	81.23

KA algoritması kullanılarak elde edilen sınıflandırma sonuçları örnek Tablo 7’de verilmiştir. Tablodan da görüleceği üzere KA algoritması 3 farklı ölçek parametresi içinde sınırlı sayıda örnek olduğu durumlarda düşük genel doğruluklar üretmiştir. Her üç ölçek parametresinde de eğitim veri seti boyutu %40 ve %70 arasında olduğu durumlarda algoritma %89 ve üzeri sınıflandırma doğruluğuna ulaşabilmiştir. KA algoritması ile en yüksek sınıflandırma doğruluğu %93,21 olarak ölçek parametresinin 59 olduğu ve bütün örneklerin %60’ının kullanıldığı durumda elde edilmiştir. Genel olarak ele alındığında, KA algoritmasının performansı eğitim veri setindeki artışla iyileşmekte ve performansında %30 seviyelerinde yükselme olduğu görülmektedir.

Tablo 7. KA algoritmasının genel doğruluk değerleri

Eğitim:Test Oranı (%)	Ölçek-20	Ölçek-32	Ölçek-59
10:90	80.81	86.34	63.28
20:80	86.47	81.85	85.39
30:70	86.32	85.51	87.54
40:60	89.55	85.47	93.11
50:50	88.84	87.53	91.26
60:40	88.80	83.43	93.21
70:30	91.95	83.59	86.48
80:20	88.56	77.83	87.85



Toplu öğrenme algoritması olan rastgele orman algoritmasının uygulanmasında ve model oluşumunda kullanıcı tarafından belirlenmesi gereken 2 parametre mevcuttur. Bunlardan ilki ormandaki ağaç sayısı ve karar ağacının her bir düğümünde rastgele seçilecek özelliklerin sayısıdır. Uygulama kapsamında RO algoritması için ağaç sayısı 500 olarak kullanılırken, rastgele seçilecek özellik sayısı kullanılan bant sayısının karekökü şeklinde hesaplanmış ve 3 olarak kullanılmıştır ( $\sqrt{10} \cong 3$ ). Her bir ölçek ve eğitim veri seti boyutu için sınıflandırma işlemleri ayrı ayrı gerçekleştirilmiş ve test veri seti için elde edilen genel doğruluk sonuçları örnek Tablo 8’de gösterilmiştir. Tablo incelendiğinde, 20 olarak seçilen ölçek parametresi için sınırlı sayıda eğitim veri seti (%10) ve mevcut örneklerin %80’nin kullanıldığı durumda RO algoritması %94,27 genel doğruluğa ulaşmıştır. Ölçek parametresi 32 olduğunda ve eğitim veri seti %60 olarak alındığında en yüksek sınıflandırma doğruluğuna (%93,24) ulaşıldığı, 59 ölçek parametresi için %30-70 eğitim verisi aralığında %93 seviyelerinde sınıflandırma doğruluğu elde edildiği görülmektedir. 20 ölçek değerinde veri setinde %30-%40 boyutlarına yaklaşıırken genel doğrulukta %3’lük azalma görülürken %80 boyutlarında tekrar yüksek doğruluğa ulaştığı tespit edilmiştir. 32 ölçek değerine bakıldığında eğitim veri setinin boyutunda ki artış genel doğruluğa %6 olarak yansıdığı görülmüştür. RO sonuçları bir bütün olarak bakıldığında eğitim veri seti boyutunun RO algoritmasının performansında etkili olduğu ve ölçek parametresi arttıkça ihtiyaç duyulan örnek sayısının arttığı görülmüştür. RO yüksek sınıflandırma doğruluğunu 59 ölçek değerinde %40 eğitim veri boyutunda elde etmiştir.

Tablo 8. RO algoritmasının genel doğruluk değerleri

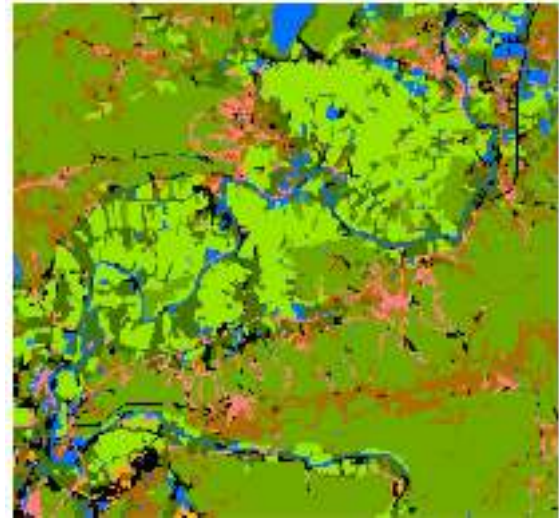
Eğitim:Test Oranı (%)	Ölçek-20	Ölçek-32	Ölçek-59
10:90	94.27	87.82	85.56
20:80	91.73	87.92	90.89
30:70	90.75	90.16	93.95
40:60	91.87	84.54	94.37
50:50	93.70	87.62	93.34
60:40	91.80	93.24	93.00
70:30	93.33	88.38	92.95
80:20	94.27	87.72	90.34

AdaBoost algoritması ile sınıflandırma işleminin gerçekleştirilmesinde kullanılacak ağaç sayısı 300 olarak seçilmiştir. AdaBoost için elde edilen sınıflandırma sonuçları örnek Tablo 9’da verilmiştir. Sonuçlar incelendiğinde 20 ölçek değeri %10 eğitim veri seti boyutunda diğer ölçeklere göre %8 artışla genel doğruluğa ulaştığı görülmüştür. Eğitim veri seti boyutu artarken genel doğruluğun %2-3 artış devam ettiği gözlenmiştir. Adaboost algoritmasında 32 değeri en düşük doğruluk seviyesinden başlayarak %40 eğitim seviyesinde %8 artışla genel doğruluğa ulaşmıştır ve eğitim veri seti %80’lere yaklaştıkça genel doğruluğun azaldığı görülmüştür. Ölçek değeri 59 değerinde başlangıçtan %10 artışla %30 eğitim veri seti boyutunda en iyi genel doğruluk sonucuna ulaştığı gözlenmiştir.

Tablo 9. Adaboost algoritmasının genel doğruluk değerleri

Eğitim:Test Oranı (%)	Ölçek-20	Ölçek-32	Ölçek-59
10:90	89.75	81.15	85.38
20:80	92.33	89.05	91.91
30:70	90.24	89.50	95.30
40:60	92.30	90.43	94.35
50:50	91.19	89.12	94.24
60:40	92.05	92.68	94.29
70:30	94.40	88.61	93.19
80:20	94.49	84.59	88.83

Uygulamada kullanılan bütün sınıflandırma algoritmalarının sonuçları karşılaştırılıp analiz edilmiştir. Analizlerin sonucuna göre en yüksek genel doğruluğu adaboost algoritması ölçek parametre değeri 59 eğitim veri boyutu 30:70 oranında ve en düşük genel doğruluğu KA algoritması 59 ölçek değerinde 10:90 eğitim veri seti boyutunda vermiştir ve bu sonuçlara göre tematik haritalar üretilmiştir. Üretilen tematik haritalar karşılaştırıldığında aralarında büyük farklılıklar olduğu göze çarpmaktadır. Genel doğruluğu düşük olan Şekil 3’deki tematik haritada yol ve şehir sınıflarının ayırt edilemez şekilde karıştığı göze çarpmaktadır. Su sınıfının diğer sınıflara karıştığı görülmektedir. Orman ve fındık sınıflarının birbirinden ayırt edilemez olduğu gözlenmektedir. Adaboost algoritması ile Şekil 4’te üretilen tematik haritada 8 sınıfın birbirinden ayırt edilebilir olduğu görülmektedir. Arazi kullanımı ve örtüsü hakkındaki bilgiye üretilen tematik haritadan çıkarımının mümkün olduğu açıkça görülmektedir.



Şekil 3. 59 ölçek ve %10 eğitim veri seti boyutu KA algoritması ile üretilen tematik harita



Şekil 4. 59 ölçek ve %30 eğitim veri seti boyutu Adaboost algoritması ile üretilen tematik harita

## 6. SONUÇLAR

Yüksek çözünürlüklü uydu görüntülerini sınıflandırmak için klasik piksel tabanlı sınıflandırma yerine nesne tabanlı sınıflandırma tercih edilmektedir. Nesne tabanlı sınıflandırma işleminin doğruluğuna doğrudan etki eden ölçek parametre değeri ve buna bağlı olarak belirlenen veri seti boyutunu belirlemek önemli bir husustur.

Bu çalışmada optimum ölçek parametre değerini belirlemek için konumsal otokorelasyon olan Moran's *I* yöntemi ve ESP2 yöntemi tercih edilmiştir. İki yöntemle toplam 3 adet ölçek parametresi belirlenmiştir. Her bir ölçek parametresi için 8 farklı veri seti boyutunda eğitim verisi toplanmıştır. Sınıflandırma işleminin gerçekleştirilmesinde, toplu öğrenme algoritmalarından Rastgele Orman (RO), Karar ağacı (KA), AdaBoost ve sınıflandırma algoritmalarından en sık kullanılan En Yakın Komşuluk (EYK) algoritmaları ile çalışılmıştır. Bu algoritmaların her bir ölçek ve eğitim veri seti boyutunda ki sonuçları karşılaştırılarak analiz edilmiştir. Çalışma sonucunda, en yüksek genel doğruluğun ESP2 aracıyla tespit edilen 59 ölçek parametre değeri ile gerçekleştirilen segmentasyon işlemi ile elde edilmiştir. Bu ölçek parametresi için tespit edilen örnek nesnelerin %30'u eğitim veri seti olarak kullanıldığında AdaBoost algoritması ile en yüksek genel doğruluğun elde edildiği görülmüştür. 3 ölçek parametresi ve 8 farklı eğitim veri setindeki genel doğruluklar kıyaslandığında genel olarak AdaBoost algoritmasının %1-2 artışla daha iyi sonuç

verdiği analiz edilmiştir. Belirlenen ölçek parametresi ve eğitim veri seti boyutlarındaki en düşük genel doğrulukların KA ile elde edildiği görülmüştür. Sınıflandırmada kullanılan algoritmaların sonuçlarının her bir ölçek ve veri seti boyutunda farklı değer vermesi bu etkenlerin genel doğruluğu etkilediği önemli bir husus olduğunu göstermektedir.

## KAYNAKLAR

Belgiu, M., Dragut, L. 2014. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. ISPRS Journal of Photogrammetry and Remote Sensing 96, pp. 67-75.

Bisht, B.S.; Kothari, B.P., 2001, Land-Cover Change Analysis of Garur Ganga Watershed Using Gis/Remote Sensing Technique, Journal of the Indian Society of Remote Sensing, 29 (3), 137-141.

Breiman, L., 2001. Random forests. In: Schapire, R.E. (Ed.), Machine Learning. Springer, Netherlands, 5-32.

Bo, S., Han, X., 2010. Parameter Selection for Segmentation in Object-Oriented Classification of Remotely Sensed Imagery, International Conference on

Intelligent Computation Technology and Automation, pp.876-879.

Çölkesen, İ., Yomralıoğlu T., ve Kavzoğlu, T., 2014 "Rotasyon Orman Algoritması İle Yüksek Çözünürlüklü Multispektral Uydu Görüntülerinin Sınıflandırılması", 5. Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu (UZALCBS'2014), 14-17 Ekim 2014, İstanbul

Çölkesen, İ., Yomralıoğlu, T., ve Kavzoğlu, T., 2015 "Objeye Tabanlı Sınıflandırmada Bölgeleme Esasına Dayalı Ölçek Parametresi Tespiti: WorldView-2 Uydu Görüntüsü Örneği", Harita Dergisi, 154(2), 9-18

Çölkesen, İ. ve Kavzoğlu T., 2017 "Kanonik Korelasyon Orman Algoritması ile Uzaktan Alınmış Görüntülerin Sınıflandırılması", Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 17(1) Özel Sayı, 102-111,

Dedeoğlu M., Başayığit L., Özyaytekin H. H., 2016 Spot 5 Uydu Görüntüsünde Objeye Tabanlı Sınıflama Yöntemleri İle Tahıl Alanlarının Parsel Haritalarının Üretimi. 6. Uzaktan Algılama-Cbs Sempozyumu , 5-7 Ekim 2016, Adana.

Drăgut, L., Tiede, D., Levick, S.R., 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. International Journal of Geographical Information Science, 24(6), pp.859-871.

Drăgut, L., Csillik, O., Eisank, C., Tiede, D., 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. ISPRS J. Photogram. Remote Sens. 88, 119-127.

- Espindola, G.M., Camara, G., Reis, I.A., Bins, L.S., Monteiro, A.M., 2006. Parameter selection for regiongrowing image segmentation algorithms using spatial autocorrelation. *International Journal of Remote Sensing*, 27(14), pp.3035-3040.
- Fotheringham, A., Brunson, C., Charlton, M., 2000. *Quantitative Geography: Perspectives on Spatial Analysis*. Sage Publications, California.
- Foody, G.M., Mathur, A., Sanchez-Hernandez, C., Boyd, D.S., 2006. Training set size requirements for the classification of a specific class. *Remote Sens. Environ.* 104 (1), 1–14.
- Kavzoğlu, T., Çölkesen, İ. (2010), Karar ağaçları ile uydu görüntülerinin sınıflandırılması: Kocaeli örneği, *Harita Teknolojileri Elektronik Dergisi*, 2, pp. 36-45.
- Kavzoğlu, T., Çölkesen, İ., ve Şahin, E.K., 2015 "Makine Öğrenme Algoritmalarının Obje Tabanlı Görüntü Sınıflandırmasındaki Etkinliğinin Araştırılması", *Türkiye Ulusal Fotogrametri ve Uzaktan Algılama Birliği VIII. Sempozyumu (TUFUAB2015)*, 21–23 Mayıs 2015, Konya.
- Kavzoğlu, T., Yıldız, M., 2015. Segmentasyonda Optimum Ölçek Parametresi Tespitinde Konumsal Otokorelasyon ve Varyansın Etkisinin Analizi, *Türkiye Ulusal Fotogrametri ve Uzaktan Algılama Birliği (TUFUAB) VIII. Teknik Sempozyumu*, 21-23 Mayıs 2015, Selçuk Üniversitesi – Konya
- Kim, M., Warner, T.A., Madden, M., Atkinson, D.S., 2011. Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. *Int. J. Remote Sens.* 32 (10), 2825–2850.
- Kim, M., Madden, M., Warner, T., 2008. Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery, In: T Blaschke, S. Lang, and G.J. Hay, (Eds.), *Object-based image analysis concepts for knowledge driven remote sensing applications*. Springer, pp. 291-307.
- Li, C., Shao, G., 2012. International Journal of Remote Object-oriented classification of land use / cover using digital aerial orthophotography, *International Journal of Remote Sensing*, 33(4), pp.922-938.
- Lillesand, T.M., Kiefer, R.W., Chipman, J.W., 2007. *Remote Sensing and Image Interpretation*. Sixth edition, New York, USA: John Wiley & Sons.
- Ma L.,cheng L., Li m.,liu Y., Ma X.,2015. Training set size, scale, and features in Geographic Object-Based Image Analysis of very high resolution unmanned aerial vehicle imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 102 pp.14-27.
- Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q. H., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, 115(5), pp. 1145-1161.
- Pal M., Mather P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4), pp. 554-565.
- Safavian S.R., Landgrebe D., 1991. A survey of decision tree classifier methodology, *IEEE Transactions on Systems Man and Cybernetics*, 21(3), pp. 660-674.
- Selva, M., B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti. 2015. Hyper-Sharpener: A First Approach on SIM-GA Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (6), 3008–3024.
- Tonbul, H., Kavzoğlu, T., (2018), "Nesne Tabanlı Görüntü Analizinde Görüntü Bölütleme Yaklaşımları ve Bölütleme Kalitesinin Analizi", *Harita Dergisi*, 84(160): 12–23
- Quinlan J.R., 1993. *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA, 302 p.