

## RASTGELE ORMAN SINIFLANDIRICISI İLE ARAZİ KULLANIM ALANLARININ BELİRLENMESİ

Ö. Akar<sup>1</sup>, O. Güngör<sup>2</sup>, A. Akar<sup>3</sup>

Karadeniz Teknik Üniversitesi, Harita Mühendisliği Bölümü, 61080, Trabzon.

<sup>1</sup>oerden@ktu.edu.tr, <sup>2</sup>ogungor@ktu.edu.tr, <sup>3</sup>aakar@ktu.edu.tr

### ÖZET

Günümüzde uydu görüntüleri, arazi kullanım alanlarının belirlenmesi, kentsel ve çevresel değişim analizleri, obje çıkarımları gibi bir çok uygulamaya altlık oluşturmaktadır. Uydu görüntülerinin sınıflandırılması da bu uygulama alanlarında kullanılan yaygın yöntemlerden birisidir. Uydu görüntüleri üzerinden en doğru ve güvenilir bilgiye ulaşmak için sınıflandırmaya yönelik, özellikle öğrenme tabanlı farklı algoritmalar geliştirilmektedir. Yaygın olarak kullanılan öğrenme tabanlı sınıflandırıcılar arasında Bagging, Boosting, Karar Ağaçları, Yapay Sinir Ağları, Destek Vektör Makineleri, K En Yakın Komşu gibi sınıflandırıcılar yer almaktadır. Son yıllarda geliştirilen Rastgele Orman (RO) yöntemi ise sunduğu yüksek sınıflandırma doğruluğu ile araştırmacıların dikkatini çekmektedir. RO, birçok sınıflandırma ağacı geliştirir. Yöntem, girdi verisinden bir objeyi sınıflandırmak için, girdi verisini ormandaki her bir ağaca yerleştirir. Her bir ağaç, bir sınıflandırma verir ve o sınıf için oylanır. Orman, ormanda yer alan tüm ağaçlardan en çok oya sahip olan sınıflandırmayı seçer. Bu çalışmada, RO sınıflandırma yöntemi kullanılarak, Trabzon ilini kapsayan 4 bantlı IKONOS uydu görüntüsü sınıflandırılmıştır. Sınıflandırma işlemi için görüntü üzerinden arazi kullanım alanları (Su, Yeşil alan, Toprak, Kentsel yapı... gibi) belirlenmiş, eğitim ve test verileri oluşturulmuştur. RO sınıflandırma algoritması ile sınıflandırılan bu görüntü, yaygın olarak kullanılan ECHO Spectral-Spatial, En Çok Benzerlik, Fisher Linear Diskriminant, En Küçük Mesafe gibi sınıflandırma yöntemleriyle karşılaştırılmıştır. RO ve diğer kullanılan yöntemlerin performanslarının karşılaştırılması sonucunda, %96.80 genel sınıflandırma doğruluğuyla RO sınıflandırma yönteminin en iyi doğruluğu sağladığı gözlemlenmiştir.

**Anahtar Sözcükler:** Sınıflandırma, Rastgele Orman, Uzaktan Algılama, IKONOS, Veri Madenciliği.

### DETERMINATION OF LAND USE AREA WITH RANDOM FOREST CLASSIFIER

#### ABSTRACT

Remote Sensing images have been widely used for many applications such as land cover/land use mapping, monitoring climate and environmental changes, pattern recognition, classification and feature extraction, etc. To acquire the most accurate and reliable information through image classification, various machine learning based algorithms have been developed. Bagging, Boosting, Decision Trees, Neural Networks, Support Vector Machines, K-Nearest Neighbor classifiers are among these algorithms. Recently developed Random Forest (RF) method has attracted the interest of many researchers with its high classification accuracy. RF grows many classification trees. To classify a new object from an input vector, the method puts the input vector down each of the trees in the forest. Each tree gives a classification, and "votes" for that class. The forest chooses the classification having the most votes over all the trees in the forest. In this study, 4-band IKONOS multispectral image taken over Trabzon is classified using RF classifier. The classes such as water, vegetation, urban fabric, etc are determined directly from the imagery by visual inspection and training and test data for each class is created. The same image is also classified using ECHO Spectral-Spatial, Maximum Likelihood, Minimum Distance and Fischer Linear Discriminant classifiers and the results are compared with the results of RF classifier. The post classification accuracy assessment results show that the RF method gives the best performance by giving 96.80% overall classification accuracy.

**Keywords:** Classification, Random Forest, Remote Sensing, IKONOS, Data Mining.

### 1. GİRİŞ

Uzaktan algılama teknikleri, çevresel ve kentsel uygulama alanlarına büyük katkı sağlamaktadır. Uydu görüntülerinin spektral, radyometrik, zamansal ve mekânsal çözünürlüklerinin giderek iyileştirilmesi yapılan uygulamaların çeşitliliğini artırmakta ve arazi kullanım alanlarının belirlenmesi, kentsel ve çevresel değişim analizleri, obje çıkarımları gibi birçok uygulama alanında avantaj sağlamaktadır. Uydu görüntülerinin sınıflandırılması da bu uygulama alanlarında kullanılan yaygın yöntemlerden birisidir. Örneğin biyolojik çeşitliliği açısından büyük önem taşıyan ve zengin bir kaynak olan tropikal ormanların izlenmesi (Christian ve Krishnaya, 2009), görüntü üzerinden kimyasal bilgilerin elde edilmesi (Cai vd., 2001), kıyı değişiminin izlenmesi (Gungor vd., 2010), kentsel gelişimin izlenmesi (Chi vd., 2009), obje çıkarımı (Zhang vd., 2007), arazi örtüsünün sınıflandırılması (Huang ve Fipps, 2006) gibi bir çok farklı alanda sınıflandırma işlemi kullanılmaktadır. Bilgi çıkartmak olarak da

adlandırılan görüntü sınıflandırma, görüntüdeki piksel değerlerini kullanarak bu bilgileri, anlamlı arazi örtüsü bilgisine dönüştürme işlemidir (Gao, 2009). Başka bir deyişle görüntü sınıflandırma, uzamsal alandaki mantıksal karar kuralları ya da çok bantlı alanlardaki istatistiksel karar kuralları kullanılarak görüntüdeki piksellerin arazi örtüsü sınıflarına göre kategorize edilmesidir. Spektral alandaki görüntü sınıflandırma, görüntü tanıma olarak da bilinir ve burada kullanılan karar kuralları uzaktan algılama verilerinin spektral değerlerini esas alır. Uzamsal görüntü tanımanın karar kurallarında ise piksellerin ya da objelerin geometrik boyutu, şekli, dokusu, deseni ön plandadır (Gao, 2009). Uzaktan algılamada görüntü sınıflandırma algoritmaları uygulama çeşitliliğine göre geliştirilmektedir. Bu algoritmalarından bazıları uydu görüntülerinin sınıflandırılmasında sıklıkla kullanılan en çok benzerlik algoritması gibi piksel tabanlıdır. Böyle yaklaşımlarda görüntünün konumsal bileşenleri ihmal edilir. Alan tabanlı sınıflandırıcılar, örneğin ECHO (Spectral-Spatial), bir obje bulma algoritması kullanmaktadır (Yıldırım vd., 2005). Hem mekânsal hem de spektral bilgiyi kullanarak sınıflandırma işlemini gerçekleştirmektedir. Alan bazlı sınıflandırma yaklaşımları objeyi parçalara ayırır ve sonrasında her bir objenin istatistiki özelliklerine göre sınıflandırma işlemini gerçekleştirir (Yıldırım vd., 2005). Yapılan çalışmalar incelendiğinde En Küçük Mesafe Sınıflandırma Yöntemi, En Çok Benzerlik Sınıflandırma Yöntemi, ECHO, Fisher Lineer Diskriminant (FLD) gibi birçok sınıflandırma yöntemlerinin yaygın olarak kullanıldığı görülmektedir. Son yıllarda uydu görüntüleri üzerinden en doğru ve güvenilir bilgiye ulaşmak için sınıflandırmaya yönelik, özellikle öğrenme tabanlı farklı algoritmalar geliştirilmektedir. Yaygın olarak kullanılan öğrenme tabanlı sınıflandırıcılar arasında Bagging, Boosting, Karar Ağaçları, Yapay Sinir Ağları, Destek Vektör Makineleri, K En Yakın Komşu gibi sınıflandırıcılar yer almaktadır. Son yıllarda geliştirilen Rastgele Orman (RO) yöntemi de yaygın olarak kullanılan sınıflandırıcılardan birisidir. Yapılan çalışmalar incelendiğinde; (Waske vd., 2007), RO, En Çok Benzerlik, karar ağaçları ve boosting yöntemlerini; (Gislason, 2004) ve (Pal, 2003), RO, boosting ve bagging yöntemlerini karşılaştırmış ve en iyi sonucu RO yöntemi ile elde etmişlerdir.

Bu çalışmada, RO sınıflandırma yöntemi kullanılarak, Trabzon ilini kapsayan 4 bantlı IKONOS uydu görüntüsü sınıflandırılmıştır. Çalışma bölgesi, Trabzon'un 41° 00' 36.22" N, 39° 43' 45.61" E; 40° 59' 53.84" N, 39° 44' 47.03" E koordinatlarını içeren Trabzon limanı ve Çömlekçi bölgesini kapsamaktadır. Bu bölge düzensiz şehirleşmenin ve gecekondu binaların yaygın olduğu, Trabzon Belediyesinin Çömlekçi Kentsel Dönüşüm Projesi kapsamında ele alınacak bir bölge olmakla dikkat çekmektedir. Görüntü üzerinden Su, Yeşil alan, Toprak, Kentsel yapı1, Kentsel yapı2, Kentsel yapı3 ve Gölge olmak üzere 7 sınıf belirlenmiştir. Bu sınıfların her biri için eğitim verileri toplanmış ve bu verilerden eğitim ve test verileri oluşturulmuştur. Çalışma alanına ait bu görüntü, yaygın olarak kullanılan ECHO Spectral-Spatial, En Çok Benzerlik, FLD, En Küçük Mesafe gibi sınıflandırma yöntemleriyle de sınıflandırılmış ve doğrulukları test edilmiştir.

## 2.SINIFLANDIRMA YÖNTEMLERİ

### 2.1. En Küçük Mesafe Sınıflandırma Yöntemi

En Küçük Mesafe Karar Kuralının temelini sınıf merkezi ile piksel arasındaki spektral mesafe oluşturur (Gao, 2009). En Küçük Mesafe Sınıflandırma yöntemi ile sınıf etiketlerini belirlemek için sınıf merkezleri ile pikseller arasındaki en küçük Öklid ya da Mahalanobis mesafeleri kullanılır. Piksel ile küme merkezi arasındaki Öklid mesafesi (1) eşitliği ile hesaplanır.

$$D_E^2 = (x_i - \mu_j)^2 \quad (1)$$

Bu eşitlikte,  $D_E$ , Öklid mesafesini;  $x_i$ ,  $i$ 'inci pikselin gözlem vektörünü;  $\mu_j$ ,  $j$ 'inci kümenin ortalama vektörünü göstermektedir.  $x_i$  vektörünün boyutu, girdi olarak kullanılan görüntünün bant sayısına eşittir. Mahalanobis mesafesi,

$$D^2 = (x_k - v_i)^T M_i^{-1} (x_k - v_i) \quad (2)$$

eşitliği ile hesaplanır. (2) eşitliğinde,  $D^2$ ,  $i$  sınıfının merkezi ile  $k$  pikseli arasındaki Mahalanobis mesafesini;  $M_i$ ,  $i$  sınıfı için varyans-kovaryans matrisini;  $v_i$ ,  $i$  sınıfı için ortalama vektörü göstermektedir. Mesafe küçüldüğünde  $i$  ve  $j$  objeleri arasındaki benzerlik artar. Her bir sınıf için sınıf

merkezi ya da spektral vektör, eğitim veri setlerinden belirlenir. Tanımsız bir piksel, kendi piksel değeri ile her bir sınıf merkezi arasındaki mesafe hesaplanarak etiketlenir.

Her bir sınıfın şekli kullanılan mesafe fonksiyonuna bağlıdır. Bu sınıflandırıcı, matematiksel olarak basittir. (Benediktsson vd., 1990) Mahalanobis mesafesinin Öklid mesafesinden daha iyi sonuç verdiğini göstermiştir. En Küçük Mesafe algoritması, En Çok Benzerlik sınıflandırma tekniğinden daha hızlı olduğu için daha cazip hale gelmiştir (Akgün, 2004). Ancak, En Küçük Mesafe algoritmasının doğruluğu, En Çok Benzerlik algoritmasının önüne geçememiştir (Tso ve Mather, 2009).

## 2.2. En Çok Benzerlik Sınıflandırma Yöntemi

Son yıllarda, Yapay Sinir Ağları, Destek Vektör Makineleri, Fuzzy, Karar Ağaçları gibi birçok sınıflandırıcı uzaktan algılama görüntülerinin sınıflandırılmasında kullanılmaktadır (Tso ve Mather, 2009). Standart En Çok Benzerlik yöntemi uzaktan algılamada en çok kullanılan sınıflandırma yöntemlerinden biridir (De Jong, vd., 2005). En Çok Benzerlik yöntemi örüntü tanımda kullanılan eğitilmiş istatistiksel bir yaklaşımdır. Her bir pikselin olasılığı hesaplanır ve piksel, olasılığı en yüksek olan sınıfa atanır. Ortalama vektörü ve kovaryans matrisi her bir sınıfı karakterize etmekte kullanılır (Tso ve Mather, 2009). En çok benzerlik yöntemi, bir pikselin her bir sınıfa ait olma olasılıklarını belirlemek için eğitim verilerinin ortalama ve varyans/kovaryans verilerini kullanır (IDRISI Klimanjaro Guide, 2003).

En Çok Benzerlik sınıflandırma yönteminin temelini olasılık yoğunluk fonksiyonu oluşturur (De Jong, vd., 2005). Bu yöntem, bir sınıfın koşullu olasılığını belirleyen Bayesion karar terorisine dayanır. (3) eşitliğinde verilen  $p(\omega_i | x)$ , aday pikselin  $i$  sınıfına ait olasılık yoğunluk fonksiyonunu temsil eder.

$$p(\omega_i | x) = \frac{p(\omega_i)p(x | \omega_i)}{p(x)} \quad (3)$$

(4) eşitliği kullanılarak  $i$  sınıfının eğitim vektörünün kovaryans ve ortalaması için  $p(x | \omega_i)$  çok değişkenli normal olasılık yoğunluk fonksiyonu hesaplanır.

$$p(x | \omega_i) = \max_j p(x | \omega_j) \quad (4)$$

En çok benzerlik diskriminant fonksiyonu;

$$g_k(x) = \ln p(\omega_i) - \frac{1}{2} \ln |C_i| - \frac{1}{2} \left[ (x - m_i)^T C_i^{-1} (x - m_i) \right] \quad (5)$$

olarak yazılabilir (Yıldırım, vd., 2005). (5) eşitliğinde,  $i$ , örnek sınıf;  $x$ , aday pikselin ölçüm vektörü;  $m_i$ ,  $i$  örnek sınıfının ortalama vektörü;  $\omega_i$  aday pikselin  $i$  sınıfına ait olma yüzdesi;  $C_i$ ,  $i$  örnek sınıfındaki piksellerin varyans-kovaryans matrisini göstermektedir.

## 2.3. ECHO Spectral-Spatial

ECHO (Extraction and Classification of Homogeneous Objects), hem mekânsal hem de spektral bilgiyi kullanarak sınıflandırma işlemini gerçekleştirmektedir. Piksel tabanlı yöntemlerde, objenin konumu, boyutu ve şekli ihmal edilip, objenin spektral özellikleri dikkate alınarak sınıflandırma gerçekleştirilir. Alan bazlı sınıflandırma yaklaşımları objeyi parçalara ayırır ve sonrasında her bir objenin istatistiksel özelliklerine göre sınıflandırma işlemini yapar (Yıldırım vd., 2005). Homojen bölgeler, görüntüde obje olarak adlandırılır. Sınıflandırmadan önce istatistiksel olarak benzer piksel dizilerini belirlemek için görüntü bölümlenmesi kullanılır. Her bir homojen dizi, bir istatistiksel örnektir. Sonrasında her bir örnek sınıflandırıcı, bu objeleri sınıflandırır. Her bir pikselin sınıflandırılması, o pikselin ve komşularının spektral özellikleri yardımıyla olur. Sınıflandırma ve bölümlenme stratejileri doğruluk, hız ve kararlılık açısından önemlidir. Bölümlenmedeki her bir obje  $K$  sınıflardan birine aittir.  $W_i$ ,  $i$  sınıfına ait olan objeleri gösterir. Objedeki her bir piksel  $q$  boyutlu rastgele bir değişkendir.  $q$  her bir pikseldeki spektral ölçüm sayısını göstermektedir. Bir  $X$  pikselinin  $q$ - olasılık değişkeni, marjini, olasılık yoğunluk fonksiyonu,  $X$ 'i içeren objenin sınıfına bağlıdır.

$$p(x \setminus W_i) = N(X; M_i; C_i) \equiv \left( |2\pi C_i| \exp\left( (x - M_i)^t C_i^{-1} (x - M_i) \right) \right)^{-1/2} \quad (6)$$

(6) Eşitliğindeki  $p(x \setminus W_i)$ ,  $i$  sınıfının olasılık yoğunluk fonksiyonunu,  $M_i$ , ortalama vektörü;  $C_i$ , kovaryans matrisini göstermektedir. Bu olasılık yoğunluk fonksiyonlarının parametrik hesabı, her bir sınıf için belirlenen eğitim veri setlerinden  $M_i$  ve  $C_i$  hesaplanarak elde edilir (Ketting ve Landgrebe, 1976).

$$\ln p(x \setminus W_i) = -\frac{1}{2} \text{tr}(C_i^{-1} S_2) + M_i^t C_i^{-1} S_1 - \frac{1}{2} n (M_i^t C_i^{-1} M_i + \ln |2\pi C_i|) \quad (7)$$

$$S_1 = \sum_{i=1}^n X_i \quad S_2 = \sum_{i=1}^n X_i X_i^t \quad M = \frac{S_1}{n} \quad C = \frac{S_2}{n - M M^t}$$

Görüntü bölümlenmede, hücre seçim kriteri;

$$Q_j(Y) = \text{tr} \left( C_j^{-1} \sum_{i=1}^m Y_i Y_i^t \right) - 2 M_j^t C_j^{-1} \sum_{i=1}^m Y_i + m M_j^t C_j^{-1} M_j \quad (8)$$

$$\begin{aligned} \ln p(Y | W_j) &= \max_i \ln P(Y | W_i) \\ &= \max_i -\frac{1}{2} (m \ln |2\pi C_i| + Q_i(Y)) \end{aligned}$$

eşitlikleriyle ifade edilir. Eşitlikte  $Y$  ise komşu gruptaki pikselleri göstermektedir. Eğer  $Q_j(Y) < c$  ise  $Y$  homojendir.  $c$  eşik değeridir. Bu durumda karar kuralı, hipotezi kabul eder. Aksi takdirde hipotez reddedilir (Ketting ve Landgrebe, 1976).

ECHO sınıflandırma yönteminin doğruluğu, az da olsa piksel tabanlı En Çok Benzerlik sınıflandırıcısından daha yüksektir. ECHO algoritmasının piksel tabanlı algoritmalarından daha karmaşık olmasına rağmen, genelde en az onlar kadar hatta onlardan daha da hızlıdır (Landgrebe, 1998). Her bir sınıflandırma büyük miktarda hesaplama gerektirir ve bu durumda sınıflandırıcı yavaşlar. ECHO sınıflandırma sayısını azaltır, maliyeti azaltırken hızda artış sağlar. (Ketting ve Landgrebe, 1976)

#### 2.4. Fisher Linear Diskriminant

FLD, veriyi boyutsal olarak küçültmek için yaygın olarak kullanılan bir yöntemdir. Bu yöntemin ana amacı, aynı anda sınıflar arası varyansı maksimum ve sınıf içi varyansı da minimum yapan bir lineer dönüşüm gerçekleştirmektir. Bu şekilde sınıfların birbirinden ayrılması mümkün olacaktır. FLD, verileri normal dağılımda kabul eder (Ball, 2006). Grafikselsel olarak tanımlanması zordur (Akgün, 2004). FDL, aşağıdaki yaklaşımı maksimum yapmayı amaçlar (Welling, 2009).

$$J(w) = (w^T S_B w) (w^T S_w w)^{-1} \quad (9)$$

(9) eşitliğinde,  $S_B$ , sınıf arası dağılım matrisini;  $S_w$ , sınıf içi dağılım matrisini göstermektedir. İki sınıf için  $S_B$ ,

$$S_B = (m_T - m_N)(m_T - m_N)^T \quad (10)$$

olarak hesaplanır. (10) eşitliğinde,  $m_T$  ve  $m_N$  hedef ve hedef olmayan sınıfların ortalama vektörlerini temsil etmektedir. FDL, hedef ve hedef olmayan dağılımları en iyi ayıran en uygun çözümü sağlamaktadır. En uygun çözüm için  $w_{OPT}$  ağırlık matrisi (11) eşitliği ile verilir.

$$w_{OPT} = S_B^{-1/2} v \quad (11)$$

$S_B^{-1/2}, S_B = S_B^{1/2} S_B^{1/2} v = S_B^{1/2} w$  'e karşılık gelir ve  $v = S_B^{1/2} w$  olarak ifade edilir.  $w$  ve  $v$  her ikisi de  $S_w^{(-1)} S_B w = \lambda w$  ifadesine karşılık gelir. (11) eşitliğinin çözümü için  $w$  eigen vektörü ve ona karşılık gelen  $\lambda$  en büyük eigen değeri kullanılır (Ball, 2006).

## 2.5. Rastgele Orman

Boosting (Shapire, 1996) ve Bagging (Breiman, 1996) ağaçların sınıflandırılmasında toplu öğrenme için çok iyi iki yöntem olarak bilinir. Bagging'de eğitim verisi kullanılarak her bir ağaç inşa edilir. Ardışık gelen ağaçlar bir öncekinden bağımsızdır ve en büyük oy tahmin için alınır. Boostingde, ardışık gelen ağaçlar bir öncekine bağımlıdır. Bir önceki öncüller tarafından yanlış tahmin edilmiş noktalar için ekstra ağırlık verilir. Sonrasında ağırlıklı oy tahmin için alınır (Liaw ve Wiener, 2002). RO, Bagging yöntemini kullanmaktadır (Breiman, 2001).

1996'da Breiman tarafından ilk amaçlanan Bagging tekniği kullanılarak her bir ağacın birbirinden bağımsız olarak eğitim verileri ile oluşturulmasıdır. Bunun yanında tahminler rastgele seçim metodu (Ho, 1998) ile oluşturulur. Ek olarak yeni verilerin tahmini için toplu öğrenme metodu süresince oluşturulan verilerden her bir değişkenin önem dereceleri ölçülür. Bu model özellikle çok sayıda öngörücü olduğunda indirgeme için yararlı olabilir. RO, çok ağaç üretmek için tekrarlanan bölümlenme ve parçalama kullanan bir toplu öğrenme metodudur (ChemModLab, 2008).

Toplu öğrenmede her bir ağacı geliştirmek için rastgele vektörler oluşturulur. Örneğin, 1996'da Breiman'ın geliştirdiği Bagging yöntemi, eğitim verisinden elde edilen örneklerin rastgele seçimiyle ağaç geliştirilmesi ilkesine dayanır. Diğer bir örnek (Dietterich 1998)'in Rastgele Split Seçimi'dir. Her bir düğümdeki splitler,  $K$  en iyi splitler arasından rastgele olarak seçilir. Breiman 1999'da, orijinal eğitim verisini rastgele hale getirerek yeni eğitim verisi oluşturmuştur. Bu yaklaşımların hepsinde,  $k$  ağaç için rastgele bir  $\theta_k$  vektörü oluşturulur. Oluşturulan  $\theta_k$  'lar birbirinden bağımsızdır. Aynı dağılıma sahip  $\theta_k$  ve eğitim verisi kullanılarak bir ağaç geliştirilir. Eğitim verisindeki örnek sayısı  $N$ 'dir. Rastgele vektörler, rastgele olarak  $N$  tane kutuya yerleştirilir. Sonrasında rastgele split seçimi yapılır. Rastgele split seçimi  $\theta$ , 1 ile  $K$  arasında yer alan bağımsız, rastgele, integer bir sayıdır.  $\theta$ 'nın boyutu ve yapısı, oluşturulacak ağacın yapısına ve kullanımına bağlıdır. Geniş sayıda ağaçlar oluşturulduktan sonra en popüler sınıf için oylanır. RO sınıflandırıcısı;

$$\{h(x, \theta_k) | k = 1, \dots\} \quad (12)$$

şeklinindedir. Burada,  $x$ , girdi verisini;  $\theta_k$ , rastgele vektörü temsil etmektedir. Her bir ağaç en popüler sınıf için bir oya atanır. Bu işlem adımlarına Rastgele Orman denir.

Breiman 2001'in Rastgele Orman yönteminde, bagging rastgele özellik seçimi ile birlikte ele alınır. Orijinal veri setinden yer değiştirmeli olarak yeni bir eğitim veri seti oluşturulur. Sonrasında, rastgele özellik seçimi kullanılarak yeni eğitim setinden bir ağaç geliştirilir. Bu geliştirilen ağaçlar budanmaz. Bagging metodunun tercih edilmesinin iki önemli nedeni vardır; birincisi, bagging işleminde rastgele özellik kullanıldığından doğruluğun artması; ikincisi, genelleştirilmiş hataları hesaplamasıdır. Bu hatalar out-of-bag (OOB) hatalarıdır (Breiman, 2001).

RO, birçok sınıflandırılmış ağaç geliştirir. Girdi verisinden yeni bir objeyi sınıflandırmak için girdi verisini ormandaki her bir ağaca yerleştirir. Her bir ağaç bir sınıflandırma verir. O sınıf için ağaç oyları belirlenir. Orman en yüksek oya sahip olan sınıflandırmayı seçer.

- 1-  $N$  sayıda orijinal veriden yer değiştirmeli olarak  $N$  sayıda rastgele eğitim verisi elde edilir.
- 2- Her bir düğüm için  $M$  toplam girdi değişkenlerinden rastgele  $m \leq M$  olacak şekilde  $m$  değişkenleri seçilir. Bu  $m$  değeri orman geliştirme süresince sabittir.
- 3- Her bir ağaç muhtemel en geniş oranda geliştirilir (Breiman ve Cutler, 2009). Sınıflandırma sırasında, kuralı durdurma ya da budama işlemleri yapılmaz (Archer, 2008). Bu durum, RO'yu diğer karar ağaçları metodlarından ayıran en önemli avantajdır (Pal, 2005).

$m$  azalınca korelasyon ve güç azalır,  $m$  artınca korelasyon ve güç artar. Bu  $m$  değeri, bulunan OOB hatalarına göre ayarlanabilir (Breiman ve Cutler, 2009). RO sınıflandırıcı ile bir ağaç üretmek için kullanıcı tarafından tanımlanan 2 parametre gereklidir. Bu parametreler, en iyi bölünmeyi belirlemek

için her bir düğümde kullanılan değişkenlerin sayısı ( $m$ ) ve geliştirilecek ağaçların sayısı ( $N$ )'dir (Pal, 2005).

Kullanıcı tarafından başlangıç  $m$  değeri rastgele seçilir sonraki  $m$ 'ler OBB hatalara göre artırılır ya da azaltılır. Bu şekilde en uygun  $m$  bulunur ve sınıflandırma duyarlılığı artar, hata azalır. RO'dan elde edilen 3 parametre vardır. Bu parametreler OOB hatası, değişken önemi (variable importance) ve yakınlık analizi (proximity analysis) (Chen, 2008).  $T$  eğitim verisinden  $T_k$  yer değiştirmeli yeni eğitim verisi üretilir. Yeni eğitim veri seti kullanılarak  $h(x, T_k)$  sınıflandırıcısı oluşturulur. Sınıflandırıcı ile çantaya atılmış tahminlerden oylama yapılır. Eğitim verisindeki her  $x$ ,  $y$  için sadece bu sınıflandırıcı ile oylama yapılır.  $T_k$ ,  $x$  ve  $y$ ' yi içermez. Bu sınıflandırıcı, OOB sınıflandırıcısı olarak da adlandırılır. Bu sınıflandırma doğruluğunun anlaşılmasına yardımcı olur (Beriman, 2001).

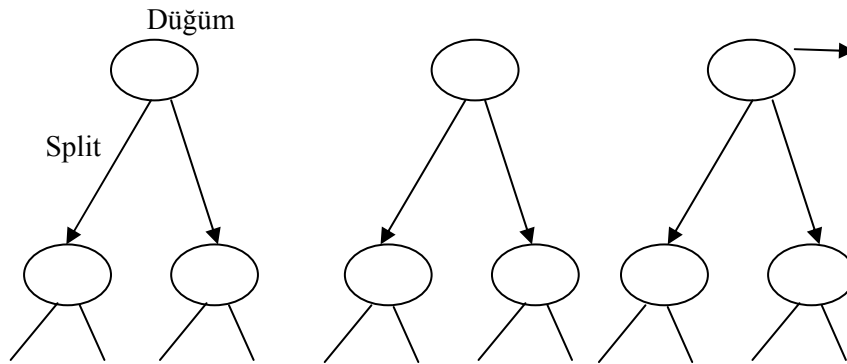
RO algoritması, OOB verisindeki verilerin yerleri değiştirildiğinde tahmin hatasının ne kadar olduğunu inceleyerek değişkenlerin önemini, etkilerini (variable importance) hesaplar (Liaw ve Wiener, 2002). Değişken önemi ölçümlerinde, kullanılan değişkenlerin ne kadar önemli olduğu değişkenlerin yerleri değiştirilerek yapılır. Değişimler sonucunda oluşan hatalar o değişkenin işlemdeki önemini ortaya koyar.

RO, budama olmadan maksimum boyutta ağaç geliştirmek için CART (Classification and Regression Tree) algoritmasını kullanmaktadır (Beriman, 2001). CART algoritmasında, bir düğümde belirli bir ölçüt uygulanarak bölünme işlemi gerçekleştirilir. Bunun için önce tüm niteliklerin var olduğu değerler göz önüne alınır ve tüm eşleşmelerden sonra iki bölünme elde edilir. Bu bölünmeler üzerinde seçme işlemi uygulanır (Özkan, 2008).

Bölünme işlemlerinde homojen sınıf dağılımına sahip düğümler tercih edilir. Düğüm homojenliğinin ölçümünde; Gini Index, Entropy, Misclassification Error, Gain Ratio Criteria gibi ölçütler kullanılmaktadır. RO yöntemi, Gini indeksini kullanmaktadır. Verilen bir  $t$  düğümü için Gini indeksi;

$$\text{GINI}(t) = 1 - \sum_j [p(j|t)]^2 \quad (13)$$

(13) Eşitliğinde,  $p(j|t)$ ,  $t$  düğümündeki  $j$  sınıfına ait bağıl olasılığı göstermektedir. En küçük gini indeksine sahip olan bölünme pozisyonu belirlenir. (Takçı, 2008). Oluşturulan eğitim verileri kullanılarak belirlenen bölünme kriterlerine göre Şekil 1'deki gibi düğümler splitlere ayrılmakta ve ağaç yapıları oluşmaktadır.



**Şekil 1.** RO sınıflandırıcısında belirlenen en uygun bölünme pozisyonlarına göre oluşturulan ağaç yapısı örneği

Kaç tane ağaç üretmek istenirse her düğüm için en iyi split belirlenerek o kadar ağaç üretilir. Belirlenen ağaçlar arasında oylama yapılır ve en iyi oyu alan ağaç bir sınıfa atanır (Liaw and Wiener, 2002). RO yönteminin sabit bir modeli, kısıtlaması, kalıbı yoktur. Kullanıcı ne kadar ağaç isterse o kadar ağaçla çalışır, hızlıdır.

K En Yakın Komşu, Destek Vektör Makineleri ve Yapay Sinir Ağları gibi çoğu öğrenme tabanlı yöntemler sınıflandırma için kullanışlıdır. Ancak bu yöntemler sınıflandırıcıya göre hangi değişkenlerin en önemli olduğuna yönelik bilgi vermezler. Lineer Diskriminant Analiz gibi diğer yöntemler de sınıflandırıcı üretirken kullanılan öncüllerde, indirgeme yapılmış öncüllere sahip tahmin uzayı gerektirir. RO yönteminde böyle bir indirgeme söz konusu değildir ve her bir adayın tahmini için değişken önem derece ölçümleri yapılır bu da sınıflandırma duyarlılığını artırmaktadır (Archer, 2008).

### 3. UYDU GÖRÜNTÜLERİNİN SINIFLANDIRILMASI VE ANALİZİ

#### 3.1. Rastgele Orman Yöntemi ile Sınıflandırma ve Analiz

RO algoritması ile görüntünün sınıflandırılması için eğitim verilerinin oluşturulması gerekmektedir. Bu işlem için ERDAS Imagine programı kullanılmıştır. Öncelikli olarak belirlenen 7 sınıfı yansıtabilecek şekilde eğitim (train) ve test verileri toplanmış ve bir eğitim dosyası oluşturulmuştur. Bu maksatla, her sınıftan eşit sayıda örnek alınarak sınıfları iyi temsil edecek toplam 6381 piksel toplanmıştır. Bu oluşturulan veriden eğitim ve test verilerinin oluşturulmasında ve RO sınıflandırma işleminde (Jaiantila, 2009)'a ait Matlab kodundan yararlanılmıştır. Çalışma kapsamında bu matlab kodu düzenlenerek, rastgele seçim ile birbirinden bağımsız olarak belirlenen 3300 piksel eğitim verisi için, 3081 piksel ise test verisi için hazırlanmıştır. Eğitim verisi görüntüyü sınıflandırmak için, test verisi ise sınıflandırma modelini test etmek için kullanılmaktadır.

RO sınıflandırıcısı için ağaç sayısı ve  $m$  değişkeni olmak üzere 2 faktör sınıflandırma doğruluğunu önemli derecede etkilemektedir. Bu parametreler kullanıcının tercihine sunulmuştur. Bunun için sınıflandırma modeline en uygun ağaç sayısı ve  $m$  değerinin tanımlanması gerekmektedir. En uygun parametrelerin bulunması işlemi için tekrarlı denemeler yapılmıştır. Öncelikle Tablo 1'de görüldüğü gibi ağaç sayısına göre değişkendeki değişimin sınıflandırmayı ne derece etkilediği test edilmiştir. Sonrasında da  $m$  değişkeni sabit tutularak uygun ağaç sayısı belirlenmiştir (Tablo2). Eğitim verisinde, görüntünün özelliğine göre Band1, Band2, Band3 ve Band4 olmak üzere toplam 4 değişken bulunmaktadır.  $M$  toplam girdi değişkenlerinden rastgele  $m$  değişkeni  $m \leq M$  olacak şekilde her bir düğüm için seçilir (Breiman ve Cutler, 2009). Varsayılan değer  $m = \sqrt{M}$  'dir (Gislason et.al., 2004). Bu nedenle  $m$  değişkenleri 1 ile 4 arasında seçilmiştir.

**Tablo 1.** RO sınıflandırıcısının 250 ağaç için  $m$  değişkenlerine göre OOB hataları ve işlem süreleri

Ağaç Sayısı	$m$ değişkeni	OOB hatası	İşlem Süresi
250	1	0.005249	377.0245 sn
250	2	0.005249	357.4112 sn
250	3	0.007874	358.3494 sn
250	4	0.005249	367.2618 sn

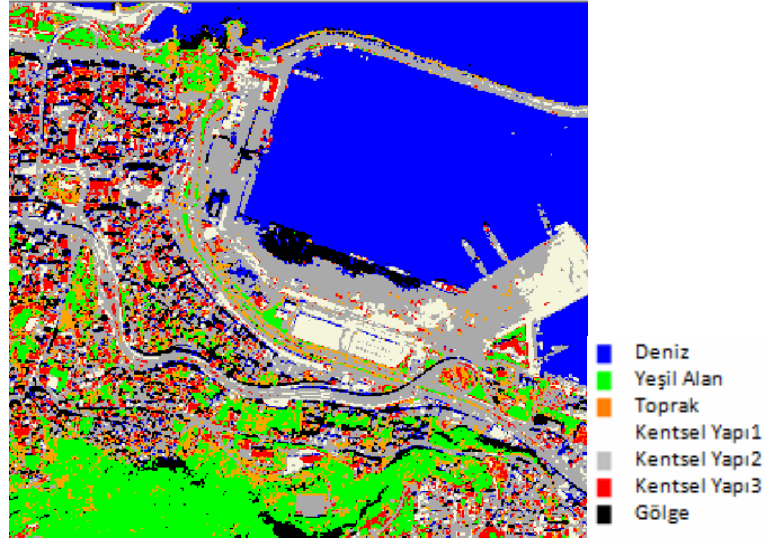
**Tablo 2.** RO sınıflandırıcısının  $m = 2$  için farklı ağaç sayılarına göre OOB hataları ve işlem süreleri

Ağaç Sayısı	$m$ değişkeni	OOB Hatası	İşlem Süresi
50	2	0.010499	143.7081 sn
100	2	0.002625	200.8782 sn
250	2	0.005249	357.4112 sn
500	2	0.005249	638.9019 sn

Seçilen parametrelere göre hem işlem süreleri hem de OOB hataları incelendiğinde RO sınıflandırması için en uygun ağaç sayısı 100 ve  $m$  parametresi de 2 olarak belirlenmiştir. Tablo 3'te görüldüğü gibi test verileri ile yapılan sınıflandırma sonucunda RO yönteminin test doğruluğu %99.55 olarak elde edilmiştir. Test verileri ile yapılan sınıflandırma doğruluğunun yüksek çıkması nedeniyle RO yöntemi tüm görüntüye uygulanmış ve Şekil 2'de verilen sınıflandırılmış görüntü elde edilmiştir.

**Tablo 3.** RO Sınıflandırma yönteminde kullanılan test verisine ait hata matrisi

	Deniz	Yeşil Alan	Toprak	Kentsel Yapı1	Kentsel Yapı2	Kentsel Yapı3	Gölge	Satır Toplamı	Producer's Accuracy	User's Accuracy
<b>Deniz</b>	2521	0	0	0	0	0	0	2521	100%	100%
<b>Yeşil Alan</b>	0	95	0	0	0	0	1	96	99%	99%
<b>Toprak</b>	0	1	26	0	1	1	0	29	84%	90%
<b>Kentsel Yapı1</b>	0	0	0	131	0	0	0	131	100%	100%
<b>Kentsel Yapı2</b>	0	0	0	0	152	0	0	152	97%	100%
<b>Kentsel Yapı3</b>	0	0	5	0	4	45	0	54	98%	83%
<b>Gölge</b>	1	0	0	0	0	0	97	98	99%	99%
<b>Sütun Toplam</b>	2522	96	31	131	157	46	98	3081		
<b>Test Doğruluğu = 99.55%</b>										
<b>Kappa = 0.9860</b>										

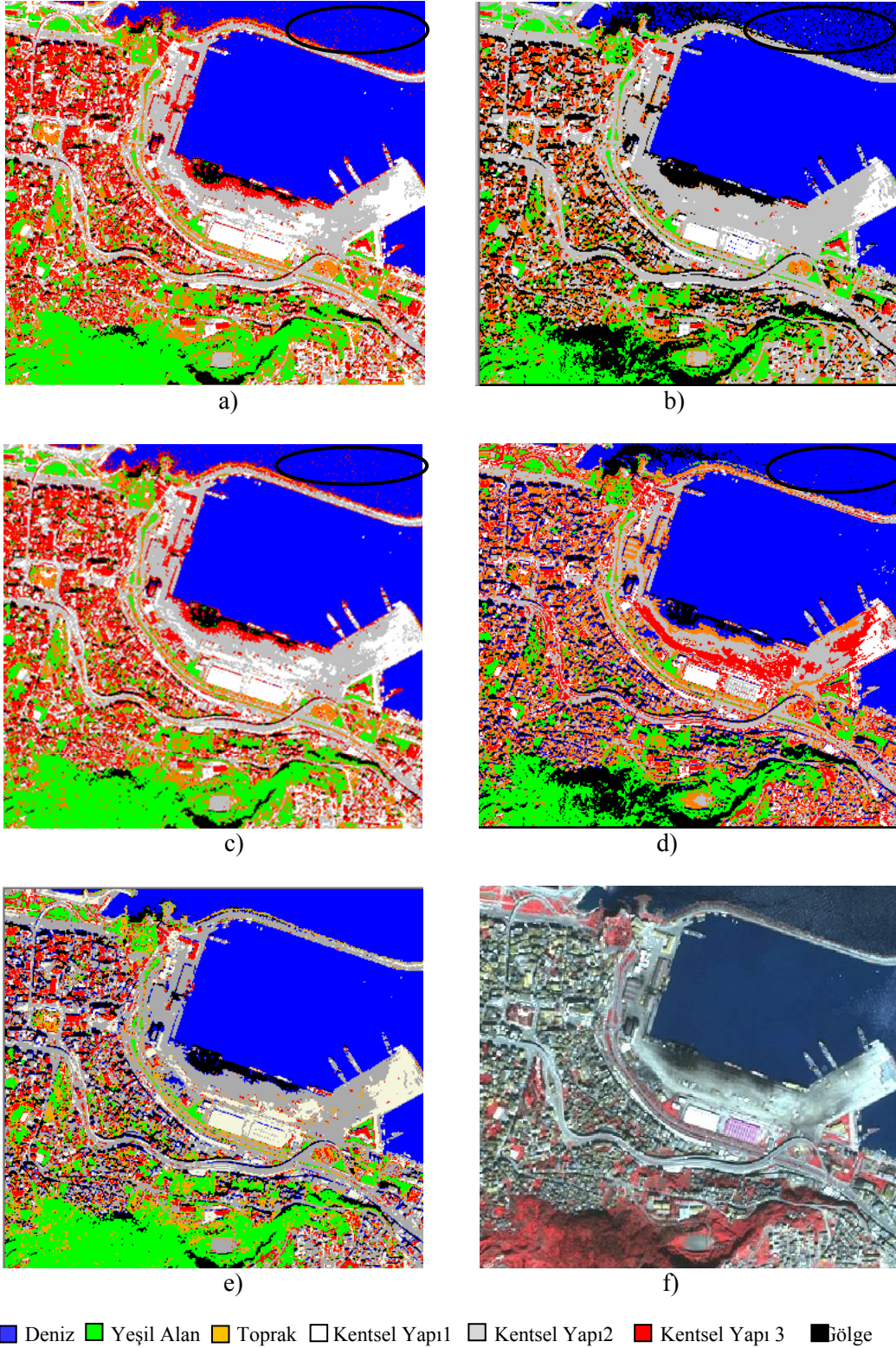
**Şekil 2.** RO ile sınıflandırılmış görüntü

### 3.2. Diğer Yöntemler ile Sınıflandırma ve Analiz

4 bantlı IKONOS uydu görüntüsü, RO haricinde popüler olarak kullanılan ECHO Spectral-Spatial, En Çok Benzerlik, FLD, En Küçük Mesafe gibi sınıflandırma yöntemleriyle sınıflandırılmıştır. Bu yöntemlerle sınıflandırma işleminde MultiSpec programı kullanılmıştır. Bu yöntemlerde de RO sınıflandırma işleminde kullanılan aynı eğitim alanları kullanılmıştır. Test verileri ile yapılan sınıflandırmalar sonucunda ECHO sınıflandırıcısı %99.3, En Çok Benzerlik %99.3, FLD %94.7 ve En Küçük Mesafe %96.7 sınıflandırma doğruluğu sağlamıştır. Daha sonra tüm görüntü bu yöntemlerle sınıflandırılmış ve sınıflandırma sonucu elde edilen görüntüler Şekil 4'te gösterilmiştir.

Şekil 4'teki sınıflandırılmış görüntüler irdelendiğinde ECHO Spectral-Spatial, En Çok Benzerlik, FLD, En Küçük Mesafe yöntemleri deniz bölgesini RO'ya göre iyi sınıflandıramamıştır. IKONOS görüntüsü konumsal olarak yüksek çözünürlüklü olduğu için denizde oluşan dalgaların sebep olduğu gölge ve parlak yansımalar hassas bir şekilde görülmektedir. Bu gölgeler ve yansımalar deniz sınıfının içinde olduğu halde farklı sınıflarla karışabilmektedir. Örneğin sınıflandırılmış görüntüler incelendiğinde gölgeler, karada binaların sebep olduğu gölgelerle ve parlak yansımaların da binalarla karıştığı gözlemlenmiştir. Kullanılan sınıflandırıcılar içinde sadece RO, bu bölgeleri doğru olarak deniz sınıfında sınıflandırmıştır. Ayrıca, görüntünün bazı bölgelerinde gölgeye ait gri değerlerin deniz sınıfındaki gri değerlere çok yakın olması sebebiyle gölge sınıfı, deniz sınıfıyla karışabilmektedir. Bu durum, En Küçük Mesafe ve RO sınıflandırıcılarında görülmüştür. Diğer sınıflandırıcılarda bu bölgeler kentsel yapı sınıflarıyla karışmıştır.



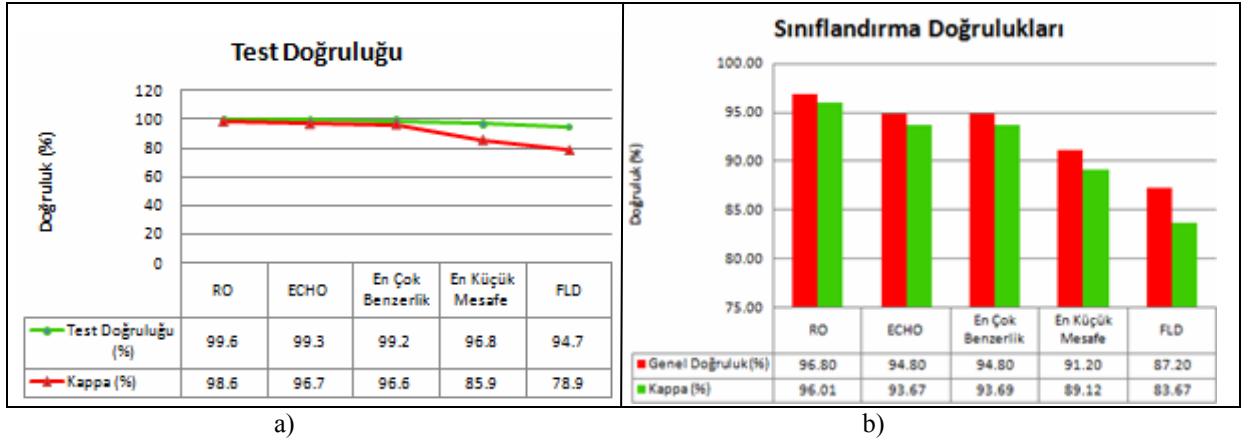


**Şekil 4.** ECHO (a), FLD (b), En Çok Benzerlik (c), En Küçük Mesafe (d), RO (e) olmak üzere 5 farklı sınıflandırma algoritmasına göre elde edilen sınıflandırılmış görüntüler ve IKONOS (f) görüntüsü

Sınıflandırma sonuçlarına göre en iyi sonucu RO, ECHO ve En Çok Benzerlik yöntemleri sağlamıştır (Şekil 5.a) Test verilerine ait hata matrisleri incelendiğinde, ECHO sınıflandırıcısı, denize ait 37 pikseli ve gölgeye ait 2 pikseli kentsel yapı3 sınıfına atamıştır. Kentsel yapı2 sınıfındaki piksellerde de, Kentsel yapı3 ve Kentsel yapı1 sınıflarıyla karışmalar olmuştur. Ayrıca Yeşil alan, Kentsel yapı2 ve Kentsel yapı3 sınıflarındaki pikseller, Toprak sınıfına atanan piksellerle karışmıştır. Aynı şekilde En Çok Benzerlik sınıflandırıcısında da Deniz ve Gölge sınıfları ile Kentsel yapı2 sınıfı arasında, Kentsel yapı2 sınıfı ile Kentsel yapı1 ve Kentsel yapı3 sınıfları arasında ve toprak sınıfı ile Kentsel

yapı<sup>2</sup> ve Kentsel yapı<sup>3</sup> sınıfları arasında karışmalar olduğu gözlenmiştir. RO sınıflandırıcısında ise Deniz, Kentsel yapı<sup>1</sup> ve Kentsel yapı<sup>2</sup> sınıfları hiçbir sınıfla karışmamıştır. Gölgedeki 1 piksel deniz sınıfına, kentsel yapı<sup>3</sup>'deki 4 piksel Kentsel yapı<sup>2</sup> sınıfına ve kentsel yapı<sup>3</sup>'deki 5 piksel de toprak sınıfına atanmıştır. Hata matrislerine göre, sınıflar en iyi RO sınıflandırıcısı ile temsil edilmiştir. Hem ECHO hem de En Çok Benzerlik sınıflandırıcıları, spektral özellikleri birbirine yakın piksellerin bulunduğu Kentsel yapı<sup>3</sup> sınıfında %56'lık, Toprak sınıfında da %86'lık doğruluk sağlamışken, RO, bu sınıflarda yaklaşık %90'lık doğruluk sağlamıştır. RO Sınıflandırıcısı birbirine yakın spektral özelliğe sahip sınıfları diğer yöntemlere göre daha iyi ve doğru şekilde sınıflandırmıştır.

Görüntülerin sınıflandırma sonrası genel doğrulukları için ERDAS Imagine programında doğruluk analizleri yapılmıştır. Analizde her sınıflandırıcı için 250 nokta kullanılmıştır. Sınıflandırma işlemleri sonucunda elde edilen genel doğruluklar Şekil 5.b ile grafiksel olarak özetlenecek olursa %96.80 sınıflandırma doğruluğu ve 0.9601 kappa değeri ile en iyi doğruluğu RO sınıflandırıcısının sağladığı görülmektedir.



Şekil 5. a) Test verileri ile yapılan sınıflandırma sonucunda elde edilen test doğrulukları, b) sınıflandırma yöntemlerinin genel doğrulukları

#### 4. SONUÇLAR

Uzaktan algılama görüntülerinden daha hassas ve doğru bilgi almak adına, sınıflandırma doğruluğu yüksek algoritmaların kullanımı giderek önem kazanmaktadır. Sınıflandırma doğruluğunu artırmaya yönelik piksel tabanlı, obje tabanlı ve öğrenme tabanlı sınıflandırıcılara yönelik birçok algoritma geliştirilmektedir. Öğrenme tabanlı sınıflandırıcılar, sağladıkları yüksek doğruluk nedeniyle yaygın olarak kullanılmaktadır. Son yıllarda geliştirilen RO, Karar ağaçları, Boosting, Bagging yöntemlerinden daha yüksek doğruluk sağlaması ve daha hızlı olması sebebiyle uzaktan algılanmış görüntülerin sınıflandırılmasına büyük avantaj sağlamaktadır.

Sınıflandırmada sıklıkla karşılaşılan sorunlardan biri de spektral olarak birbirine çok yakın olan sınıfların ve piksellerin yanlış sınıflara atanmasıdır. Özellikle yüksek çözünürlüklü görüntülerde birbirine yakın spektral özelliklere sahip piksellere rastlanmaktadır. Örneğin bu çalışmada kullanılan yüksek çözünürlüklü IKONOS görüntüsünde, deniz dalgalarının bina ve gölge sınıflarına ve gölgelerin de deniz sınıfına yakın spektral özellikte olması, bu problemi açıkça ortaya koymaktadır. Bu çalışmada yaygın olarak kullanılan alan bazlı sınıflandırıcı olan ECHO, piksel tabanlı sınıflandırıcı olan En Küçük Mesafe, piksel ve istatistiksel tabanlı En Çok Benzerlik, FLD ve RO sınıflandırıcıları arasından, bu tür problemler için çözüm olabilecek en doğru ve en hızlı sınıflandırıcının bulunması amaçlanmıştır. Bunu için çalışma alanına ait eğitim verileri oluşturulmuş, bu verilere göre farklı sınıflandırıcılar kullanılarak sınıflandırmalar yapılmıştır. Ayrıca sınıflandırılmış görüntüler üzerinden rastgele seçilen 250 nokta ile doğruluk analizleri gerçekleştirilmiştir. Bunların sonucunda, RO sınıflandırıcısı ile elde edilen görüntüde test doğruluğu %99.5, sınıflandırma sonrası genel doğruluk ise %96.80 olarak belirlenmiştir. ECHO sınıflandırıcısı %94.80, En Çok Benzerlik %94.80, FLD %87.20 ve En Küçük Mesafe %91.20 sınıflandırma sonrası genel doğruluk değerlerini vermiştir. Elde edilen sonuçlar değerlendirildiğinde, RO sınıflandırıcısının, diğer sınıflandırıcıların birbirinden ayıramadığı spektral olarak birbirine yakın olan nesnelere gerçeğe en yakın olarak sınıflandırdığı ve bu sınıflar için en hızlı ve en doğru sınıflandırma sonucunu verdiği gözlemlenmiştir.

## KAYNAKLAR

- Akgün A., Eronat A.H. , Türk N.**, 2004, Comparing different satellite image classification methods: an application in Ayvalık district, Western Turkey. <http://www.isprs.org/proceedings/XXXV/congress/comm4/papers/505.pdf> (05.06.2010).
- Archer, K. J.**, 2008, Empirical characterization of random forest variable importance measure, computational statistical data analysis, *Computational Statistics & Data Analysis*, 52(4), 2249-2260.
- Ball, J., Bruce, L.M.**, 2006, Accuracy analysis of hyperspectral imagery classification using level sets, ASPRS 2006 Annual Conference, Reno, Nevada, May 1-5, 2006.
- Benediktsson, J. A., P. H. Swain, O. K. Esroy**, 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data, *IEEE Transactions on Geoscience and Remote Sensing* 28(4), 540–552.
- Breiman, L., Cutler, A.**, 2009, Random Forest, [http://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_home.htm](http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm) (15.11.2009).
- Breiman, L.**, 2001, Random Forests, *Machine learning*, 2001 Kluwer Academic Publishers, 45(1), 5-32.
- Cai, T.T., Zhang, D., Ben-Amotz D.**, 2001, Enhanced chemical classification of raman images using multiresolution wavelet transformation, *Applied Spectroscopy*, 55(9), 1124-1130.
- ChemModLab**, 2008, Random Forest, ChemModLab Documentation, <http://eccr.stat.ncsu.edu/ChemModLab/RandomForest.pdf> (23.12.2009).
- Chen, C. H.**, 2008, *Image Processing For Remote Sensing*, CRS Press, Taylor&Francis Group, USA.
- Chi, M.V., Thi, L.P, Si, S.T**, 2009, Monitoring urban space expansion using remote sensing data in Ha Long City, Quang Ninh Province in Vietnam, 7th FIG Regional Conference Spatial Data Serving People: Land Governance and the Environment – Building the Capacity Hanoi, Vietnam.
- Christian B., Krishnappa N. S. R**, 2009, Classification of tropical trees growing in a sanctuary using Hyperion (EO-1) and SAM algorithm, *Current Science*, 96(12), 1601-1607.
- De Jong, S. M, Van Der Meer, F. D**, 2005, *Remote Sensing Image Analysis: Including the Spatial Domain*, Springer Science, Volume 5, United States of America.
- Gao, J.**, 2009, *Digital Analysis of Remotely Sensed Imagery*, The McGraw-Hill Companies, USA.
- Gislason, P.O., Benediktsson, J.A, Sveinsson, J.R.**, 2004, Random forest classification of multi-source remote sensing and geographic data, *IEEE Geoscience and Remote Sensing Symposium*, Vol.2, 1049–1052.
- Gungor O., Boz Y., Gokalp E., Comert C., Akar A.**, 2010, Fusion of low and high resolution satellite images to monitor changes on coastal zones, *Scientific Research and Essays*, 5(7), 654-662.
- Huang, Y., Fipps, G.**, 2006, Landsat satellite multi-spectral image classification of land cover change for GIS-based urbanization analysis in irrigation districts: evaluation in Low Rio Grande Valley1, <http://idea.tamu.edu/documents/LandsatReport06Aug.pdf> ( 21.02.2010).
- IDRISI Klimanjaro**, 2003, *Guide to GIS and Image Processing*, Idrisi Production, Clark Labs, USA, pp.57-82.
- Jaiantila, A.**, 2009, Classification Random Forest, [http://code.google.com/p/randomforest-matlab/downloads/detail?name=RF\\_MexStandalone-v0.02.zip&can=2&q=](http://code.google.com/p/randomforest-matlab/downloads/detail?name=RF_MexStandalone-v0.02.zip&can=2&q=) (25.08.2009).
- Ketting R.L., Landgrebe, A.**, 1976, Classification of multispectral image data by extraction and classification of homogeneous objects, *IEEE Transactions On Geoscience Electronics*, 14(1), 19-26.
- Landgrebe, D.**, 1998, *Multispectral data analysis: a signal theory perspective*, Indiana: School Of Electrical And Computer Engineering, Purdue University, West Lafayette. [https://engineering.purdue.edu/~biehl/MultiSpec/Signal\\_Theory.pdf](https://engineering.purdue.edu/~biehl/MultiSpec/Signal_Theory.pdf) (03.07.2010)
- Liaw, A., Wiener, M.**, 2002, Classification and Regression By Random Forest, *R News*, Vol.2/3, December.
- Özkan, Y.**, 2008, *Veri Madenciliği Yöntemleri*, Papatya Yayıncılık, İstanbul.
- Pal, M.**, 2005, Random Forest Classifier For Remote Sensing Classification, *International Journal of Remote Sensing*, 26(1) , 217-222
- Pal, M.**, 2003, Random Forest For Land Cover Classification, *IEEE Geoscience and Remote Sensing Symposium*, Vol.6, 3510–3512.
- Takçı, H.**, 2008, *Veri Madenciliği Dersi Notları*, Gebze Yüksek Teknoloji Enstitüsü, Dr. Hidayet Takçı.
- Tso, B., Mather P.M.**, 2009. *Classification Methods For Remotely Sensed Data*, Second Editon, Taylor & Francis Group, London.
- Waske, B., Heinzl, V., Braun M., Menz, G.**, 2007, Random forest for classifying multi-temporal SAR data, *Proc. Envisat Symposium 2007*, Montreux, Switzerland, 23–27 April 2007 (ESA SP-636, July 2007), <http://envisat.esa.int/envisatsymposium/proceedings/sessions/3D3/461589wa.pdf> (13.03.2010).
- Welling, M.**, 2009, Fisher Linear Discriminant Analysis, Lecture notes, University of Toronto, Canada. <http://www.ics.uci.edu/~welling/teaching/273ASpring09/Fisher-LDA.pdf>.
- Yıldırım, İ., Ersoy, O.K., Yazgan B.**, 2005, Improvement of classification accuracy in remote sensing using morphological filter, *Advances in Space Research*, 36 1003–1006.
- Zhang W., Xue X., Sun Z., Guo Y., Chi M., Lu H.**, 2007, Efficient feature extraction for image classification, *IEEE 11th International Conference on Computer Vision*, pp.1-8