

# İYİLEŞTİRİLMİŞ KARAR AĞAÇLARI İLE UYDU GÖRÜNTÜLERİNİN SINIFLANDIRILMASI

İ. Çölkesen<sup>1</sup>, T. Kavzoğlu<sup>2</sup>  
GYTE, Gebze Yüksek Teknoloji Enstitüsü, Jeodezi ve Fotogrametri Mühendisliği Bölümü,  
<sup>1</sup>[icolkesen@gyte.edu.tr](mailto:icolkesen@gyte.edu.tr), <sup>2</sup>[kavzoglu@gyte.edu.tr](mailto:kavzoglu@gyte.edu.tr)

## ÖZET

Sahip olduğu önemli avantajlarla uzaktan algılama teknolojileri, farklı disiplinler tarafından global ve yerel ölçekte gerçekleştirilen birçok uygulamada başvurulan önemli bir kaynak durumundadır. Bu teknolojilerin temel ürünü olan uydu görüntüleri üzerinden yeryüzüne ait önemli bilgilerin elde edilmesinde sınıflandırma tekniklerinden yararlanır. Bu amaçla günümüze kadar farklı sınıflandırma algoritmaları geliştirilmiş ve çeşitli derecelerde başarılar elde edilmiştir. Uzaktan algılanan görüntülerin sınıflandırılmasında son yıllarda kullanılmaya başlayan karar ağaçları akış şemalarına benzeyen yapılarıyla birçok alanda başarıyla kullanılan bir kontrollü sınıflandırma yöntemidir. Literatürde sınıflandırma yöntemlerinin performansını arttırmaya yönelik geliştirilen çeşitli iyileştirme yöntemlerinden bahsedilmektedir. Bunlar arasında hızlandırma, torbalama ve çoklu hızlandırma algoritmaları en çok bilinen ve kullanılan iyileştirme yöntemlerindedir. Bu çalışmada, sınıflandırmaya esas veri seti olarak Gebze ilçesini kapsayan 2002 tarihli Terra ASTER uydu görüntüsü kullanılarak karar ağacı sınıflandırma metodunun performansı değişik iyileştirme metotları da kullanılarak ayrıntılı olarak analiz edilmiştir. Son olarak yöntemin sınıflandırma performansı en çok benzerlik sınıflandırma yönteminin performansı ile karşılaştırılmıştır. Elde edilen sonuçlar, iyileştirme algoritmalarıyla güçlendirilmiş karar ağacı sınıflandırıcılarının uzaktan algılanmış görüntülerin sınıflandırılmasında yüksek doğruluklu sonuçlar ürettiğini ve bu nedenle iyi bir alternatif olabileceğini göstermektedir.

Anahtar Sözcükler: Sınıflandırma, En Çok Benzerlik, Karar Ağaçları, Hızlandırma, Torbalama

## ABSTRACT

### CLASSIFICATION OF SATALLITE IMAGES USING IMPROVED DECISION TREES

Remote sensing technologies having significant advantages are important sources for many applications in many fields at global and local scales. Classification techniques have been used to obtain crucial information about the Earth's surface from the satellite images that are primary product of remote sensing technologies. For this purpose, many algorithms have been developed and applied in the literature for the classification of remotely sensed image. Decision trees whose structures are similar to flow-charts are a group of classification algorithms that have been recently used in the classification of remote sensing images and successfully applied to many applications. In the literature, various enhancement methods are developed and discussed to increase the performance of classification methods. Among these methods boosting, bagging and multi-boosting are the most commonly known and applied algorithms. In this research, Terra ASTER image acquired in 2002 covering the Gebze district of Turkey were used for the main data set and performance of decision tree classifier was analyzed together with three enhancement methods. Later, classification performance of decision trees was compared with the popular classification method namely maximum likelihood classifier. Results produced in this research show that decision tree classifier with enhancement algorithms can produce high classification accuracies for the remotely sensed images; thus, they appear to be a good alternative to conventional classification techniques.

Keywords: Classification, Maximum Likelihood, Decision Trees, Boosting, Bagging

## 1. GİRİŞ

Küresel ve yerel ölçekte yapılan birçok çalışmada yeryüzüne ait bilgilerin hızlı ve güvenilir bir şekilde elde edilmesi önemli bir adımdır. Bu bilgilerin elde edilmesinde kullanılan diğer yöntemlerin yanında uzaktan algılama teknolojileri sahip olduğu özellikler ile önemli avantajlara sahiptir. Yeryüzünün değişik zaman aralıklarında, farklı konumsal ve spektral çözünürlüklerde görüntülenebilmesi bu teknolojilerin en önemli avantajları olarak ifade edilebilir. Tüm bu özellikleriyle uzaktan algılanan görüntüler yeryüzüne ait bilgilerin elde edilmesi noktasında önemli bir kaynak durumundadır. Bu görüntüler üzerinden arazi örtüsü veya kullanım şekilleri gibi yeryüzüne ait önemli bilgilerin elde edilmesinde kullanılan en önemli yöntem uydu görüntülerinin sınıflandırılmasıdır.

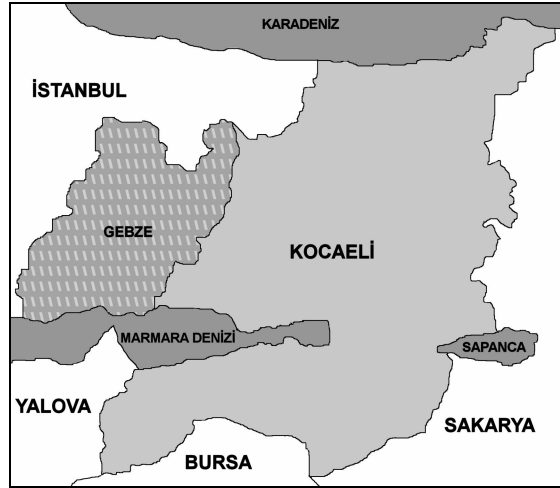
Uzaktan algılanmış görüntülerin sınıflandırılması işlemi genel olarak görüntüdeki her bir pikselin yansıtma ve parlaklık değerlerinden faydalanılarak bu pikselleri belirli arazi örtüsü sınıflarından birine atanması olarak tanımlanabilir (Huang vd., 2001). Başka bir ifadeyle, görüntüdeki pikselleri arazide karşılık geldikleri sınıflar veya temalar içine atamak veya yerleştirmektir. Richards (1993) görüntü sınıflandırmasını çeşitli sınıflar ile objelere ait spektral imzalar arasındaki ilişkiyi kurarak görüntüler üzerinden tematik bilgilerin çıkarıldığı bir görüntü işleme bileşeni olduğunu ifade etmiştir.

Literatürde uzaktan algılanmış görüntülerin sınıflandırılmasında kullanılmak üzere birçok sınıflandırma algoritması geliştirilmiştir (Townshend, 1992; Hall vd., 1995; Lu and Weng, 2007). Bu algoritmalar arasında günümüzde geçerliliği kabul edilen ve yaygın bir kullanıma sahip olan en çok benzerlik (EÇB) sınıflandırıcısıdır. Son yıllarda uzaktan algılanmış görüntülerin sınıflandırılması üzerine yapılan çalışmalarda kullanılan karar ağaçları (KA) hesaplama açısından hızlı ve istatistiksel kabullere dayanmayan diğer bir sınıflandırma yöntemidir (Huang vd., 2002; Pal ve Mather, 2003). Literatürde özellikle karar ağaçlarının sınıflandırma performansını arttırmaya yönelik geliştirilen çeşitli iyileştirme yöntemleri mevcuttur. Bunlar arasında birden çok sınıflandırıcının birleştirilmesi ile tek bir sınıflandırıcının elde edilmesi sık kullanılan bir iyileştirme yaklaşımıdır. Birleştirme sonucu elde edilen sınıflandırıcı ile yapılan sınıflandırma doğruluğunun genel olarak her bir sınıflandırıcının tekil olarak kullanılmasından daha iyi olduğu ifade edilmektedir (Opitz ve Maclin, 1999). Sınıflandırıcıların birleştirilmesinde hızlandırma, torbalama ve çoklu hızlandırma metotları en yaygın kullanılanlar arasındadır (Schapire, 1999; Breiman, 1996; Webb, 2000).

Bu çalışmada, karar ağaçları ile 2002 tarihli Terra ASTER uydu görüntüsü kullanılarak Gebze ve çevresindeki arazi kullanımının tespiti amaçlanmıştır. Karar ağaçlarının tekil kullanımının yanında hızlandırma (Adaboost), torbalama ve çoklu hızlandırma iyileştirme metotları ile yöntemin sınıflandırma performansındaki değişimler belirlenmiştir. Son olarak KA'nın sınıflandırma performansı popüler sınıflandırma yöntemlerinde en çok benzerlik sınıflandırıcısı ile karşılaştırılarak analiz edilmiştir.

## 2. ÇALIŞMA ALANI

Gebze, Marmara bölgesinde Kocaeli ilinin kuzeybatısında, İstanbul'un 45 kilometre doğusunda yer alan bir ilçedir. Türkiye sanayisinin %15'ini, Kocaeli'nin endüstrisinin büyük bölümün barındıran Gebze, Marmara bölgesinin en büyük ikinci ilçesidir. Yaklaşık olarak yüz ölçümü 772 km<sup>2</sup> olan ilçe, coğrafi konumu itibariyle 40° 45' 08'' ile 41° 02' 38'' kuzey enlemleriyle 29° 19' 56'' ile 29° 45' 14'' doğu boylamları arasında yer almaktadır (Şekil 1). Doğusunda İzmit, batısında İstanbul, kuzeyinde Şile, güneyinde (İzmit Körfezi) ile sınırlanan Gebze, yayla görünümünde bir düzlüğün üzerinde kurulmuştur (URL 1).



Şekil 1: Çalışma alanı (Gebze) haritası

## 3. MATERYAL VE YÖNTEM

Bu çalışmada, arazi örtüsünün sınıflandırılması için çalışma alanını kapsayan 2002 yılında kaydedilmiş Terra ASTER görüntüsü kullanılmıştır. Uydu görüntüsünün geometrik düzeltilmesinde bölgeye ait 1/25.000 ölçekli haritalardan yararlanılmıştır. Sayısal olarak bilgisayar ortamına aktarılan 1/25.000 ölçekli haritalar ERDAS/Imagine yazılımı ile UTM koordinat sisteminde tanımlanarak haritalar mozaik haline getirilmiştir. Ayrıca, uydu görüntüsünün sınıflandırılması amacıyla eğitim alanlarının belirlenmesinde farklı tarihlerde çekilmiş hava fotoğrafları (1996, 1999 ve 2003) ve Gebze orman bölge şefliğine ait meşcere haritalarından yararlanılmıştır. Bunların yanı sıra arazide görsel olarak da alanlara ait konumlar el GPS aletleriyle tespit edilmiştir.

Uydu görüntülerinin farklı veri gruplarıyla beraber değerlendirilebilmesi için aynı koordinat sisteminde olmaları gerekmektedir (Mather, 1987). Uydu görüntülerinin geometrik dönüşümü için, UTM projeksiyon sisteminde sayısal ortama aktarılmış topoğrafik haritalardan faydalanılmıştır. Koordinat dönüşümünde yer kontrol noktalarının tüm çalışma alanına homojen olarak dağılmasına dikkat edilmiştir. Uydu görüntülerine birinci derece Afin dönüşümü kullanılarak UTM projeksiyon sisteminde yaklaşık 0.5 piksel karesel ortalama hata ile geometrik düzeltme getirilmiştir. Uydu görüntülerinin yeniden örneklenmesinde orijinal değerleri koruduğundan en yakın komşu yöntemi tercih

edilmiştir. Yapılan ön çalışmalar neticesinde çalışma alanında altı temel sınıfın mevcut olduğu belirlenmiştir. Belirlenen bu sınıflar deniz, yerleşim, toprak-taş, bozkır, iğne yapraklı orman ve geniş yapraklı orman sınıflarıdır.

#### 4. SINIFLANDIRILMA YÖNTEMLERİ

Ham halde elde edilen uydu görüntülerinden yeryüzüne ait bilgilerin elde edilmesi için çeşitli istatistiksel analizler ve istatistiksel yorumlama teknikleri kullanılmaktadır. Verileri bilgiye dönüştürebilmek için en yaygın yöntem görüntü sınıflandırma yöntemidir. Uzaktan algılamada sınıflandırma, cisimlerin farklı spektral yansıtma değerleri esasına dayanarak orijinal görüntüdeki her görüntü elemanını ait olduğu özellik grubuna ayırma işlemidir. Sınıflandırmada amaç, yeryüzü üzerinde aynı spektral özellikleri taşıyan nesnelere gruplandırılmasıdır. Sınıflandırma sonucu elde edilen görüntü tematik harita olarak adlandırılmaktadır. Sınıflandırma işleminde genel olarak kontrollü ve kontrolsüz sınıflandırma olarak adlandırılan iki yaklaşım vardır. Bu çalışmada, belirlenen altı sınıfın çalışma alanını kapsayan uydu görüntüsü kullanılarak sınıflandırılması işleminde kontrollü sınıflandırma teknikleri olan karar ağaçları ve en çok benzerlik yöntemleri kullanılmıştır. Hızlandırma, torbalama ve çoklu hızlandırma olarak bilinen iyileştirme yöntemleri ile karar ağaçlarının sınıflandırma performansındaki değişimler belirlenmiştir.

##### 4.1 En Çok Benzerlik Sınıflandırma Yöntemi

En çok benzerlik yöntemi literatürde en yaygın olarak kullanılan etkili bir sınıflandırma yöntemidir. Bu yöntemde ortalama değer, varyans ve kovaryans gibi istatistiksel değerlerin tümü dikkate alınır. Kontrol alanlarını oluşturan sınıflar için olasılık fonksiyonları hesaplanmakta ve buna göre her bir pikselin hangi sınıfa daha yakın olduğuna karar verilmektedir. Bir pikselin hangi sınıfa ait olduğu her bir sınıfa ait olma olasılıklarının hesabından sonra en yüksek olasılıklı grubu atama şeklinde yapılır. Bu noktada kullanıcı tarafından tespit edilecek bir eşik değeri sınıflandırılacak pikselin belirlenen sınıflardan veya bu sınıfların dışında bir sınıftan olduğunun tespitinde kullanılır. Bu yöntemde, sınıf kontrol verilerini oluşturan noktalar kümesindeki dağılımın, Normal dağılım olduğu kabul edilir. Sınıfların ilk olasılıkları hakkında bilgi mevcut değilse, hepsi eşit olasılıklı olarak kabul edilir. Yöntem, pikselleri sadece parlaklık değerlerine göre değil, her sınıf için ayırım oluşturacak varyans-kovaryans matris değerine göre oluşturur. Böylece örnek piksellerin özellik uzayındaki dağılımları da dikkate alınmış olur (Mather, 1987). En çok benzerlik yönteminde kullanılan formül Eşitlik (1) ile ifade edilebilir.

$$D = \ln(a_c) - [0.5 * \ln(|Cov_c|)] - \left[ 0.5 * (X - M_c)^T * (Cov_c^{-1}) * (X - M_c) \right] \quad (1)$$

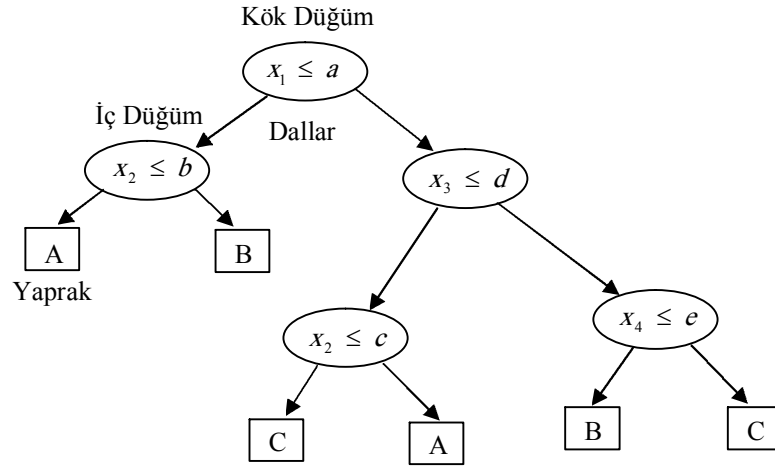
Bu eşitlikte;  $D$ , uzaklık ağırlıklı olasılık değerini;  $C$ , örnek bir sınıfı;  $X$  aday pikselin ölçüm vektörünü;  $M_c$ ,  $C$  örnek sınıfının ortalama vektörünü;  $a_c$ , aday pikselin  $C$  sınıfına ait olma yüzdesini,  $Cov_c$ ,  $C$  örnek sınıfındaki piksellerin varyans-kovaryans matrisini göstermektedir.

##### 4.2 Karar Ağaçları

Karar ağaçları son yıllarda literatürde yaygın kullanımı olan sınıflandırma ve örüntü tanımlama algoritmalarından birisidir. Bu yöntemin yaygın olarak kullanımının en önemli nedeni ağaç yapılarının oluşturulmasında kullanılan kuralların anlaşılabilir ve sade olmasıdır. Karar ağaçları (KA) arazi örtüsünün sınıflandırılmasında veya sınıf etiketlerinin belirlenmesinde çok aşamalı veya ardışık bir yaklaşım kullanılmaktadır. Yöntem çok karışık bir sınıflandırma problemini çok aşamalı bir hale getirerek basit bir karar verme işlemi gerçekleştirir (Safavian ve Landgrebe, 1991). Ağaç oluşumunun her bir aşamasında kullanılan değişkenlerin sayısına bağlı olarak tek değişkenli veya çok değişkenli karar ağacı yapıları mevcuttur (Friedl ve Brodley 1997). Tek değişkenli karar ağaçları genellikle arazi örtüsü sınıflandırmalarında kullanılan ağaç yapılarıdır (DeFries vd., 1998, Hansen et al. 2000). Çok değişkenli karar ağaçları tek değişkenli karar ağaçlarına göre daha hassas olmalarına rağmen, daha karmaşık algoritmalar içermektedirler. Bu karmaşık algoritmalar ise çok farklı faktörlerden etkilenmektedir (Brodley ve Utgoff, 1995).

Bir karar ağacının temel yapısı Şekil 2'de görüldüğü üzere düğüm, dal ve yaprak olarak adlandırılan üç temel kısımdan oluşur. Bu ağaç yapısında her bir öznelik bir düğüm tarafından temsil edilir. Dallar ve yapraklar ağaç yapısının diğer elemanlarıdır. Ağaçta en son kısım yaprak en üst kısım ise kök olarak adlandırılır. Kök ve yapraklar arasında kalan kısımlar ise dal olarak ifade edilir (Quinlan, 1993). Veri seti ağaç tarafından belirlenen karar iskeletine göre aşağı doğru hareket ettirilerek ve sıralı olarak bir yaprağa ulaşılan kadar alt bölümlere ayrılarak sınıflandırılır. Eğitim verileri ve bu verilere ait öznelik bilgilerinden yararlanılarak bir karar ağacı yapısı oluşturulmasında temel prensip verilere ilişkin bir dizi sorular sorulması ve elde edilen cevaplar doğrultusunda hareket edilerek en kısa sürede sonuca gidilmesi olarak ifade edilebilir. Bu şekilde karar ağacı sorulara aldığı cevapları toplayarak karar kuralları oluşturur. Bu noktada belirtilmesi gereken husus yaprak sayısı kadar kural oluşacağıdır (Pal, 2007). Oluşturulan ağacın işlevselliğinin başka bir ifadeyle yeni bir veri seti için genelleme kabiliyetinin belirlenmesi için test veri seti kullanılır. Eğitim veri seti ile oluşturulan ağaç yapısına yeni gelen bir test verisi, ağacın kökünden girer. Kökte test edilen bu yeni veri test sonucuna

göre bir alt düğüme gönderilir. Ağacın belirli bir yaprağına gelene kadar bu işleme devam edilir. Şekil 2’de üç sınıfa ait dört boyutlu öznelik değerlerinden oluşan basit bir ağaç yapısı görülmektedir. Şekilde  $x_i$  öznelik değerlerini;  $a, b, c, d$  ve  $e$  değerleri dallanma için esas olan eşik değerlerini ve A, B ve C ise sınıf etiketlerini göstermektedir.



Şekil 2: Dört boyutlu özellik uzayına sahip üç sınıftan oluşan bir karar ağacı yapısı

Karar ağaçlarının oluşturulmasındaki en önemli adım ağaçtaki dallanmanın hangi kritere veya kıstasa göre yapılacağı ya da hangi öznelik değerlerine göre ağaç yapısının oluşturulacağıdır. Tek değişkenli karar ağacı oluşturulmasının temelini oluşturan Hunt algoritmasında bu kriterlerin rastgele seçilmesi algoritmanın en önemli eksikliği idi (Quinlan, 1993). Bu eksikliğin giderilmesi ve yöntemin geliştirilmesi amacıyla çeşitli algoritmalar geliştirilmiştir. Bunlardan en önemlisi Quinlan (1987) tarafından ortaya konulan ID3 algoritmasıdır. Bu algoritma sayısal olmayan verilerin sınıflandırılması için geliştirilmiş entropiye dayalı bir algoritmadır. ID3 algoritmasının geliştirilmiş hali olan C4.5 algoritması ise bölünme bilgisi kavramı ile bilgi kazancından yararlanarak hesaplanan kazanç oranı yaklaşımını kullanmaktadır.

Karar ağacı oluşturulmasında diğer önemli bir husus oluşturulan ağacın budanması işlemidir. Karar ağacı sınıflandırıcısı eğitim verisini sadece tek bir sınıf içeren alt kümelerle böler. Bu işlem sonucunda çok geniş ve karmaşık bir ağaç ortaya çıkar. Karar ağaçlarında, bir alt ağacı atarak yerine bir yaprak yerleştirmek söz konusu olabilir. Karar ağaçları bir veya daha fazla alt ağacın kaldırılması ile sadeleştirilir ve bu alt ağaçların yerine yapraklar ile ifade edilir. Bu şekilde gerçekleştirilen işleme karar ağacının budanması adı verilir (Özkan, 2008; Pal, 2007). Budama ile karar ağacının sınıflandırma doğruluğunu etkilemeyen veya katkısı olmayan kısımlarının çıkarılması işlemi gerçekleştirilir. Böylece daha az karmaşık ve daha anlaşılabilir bir ağaç elde edilir. Daha basit bir ağaç oluşturulması için karar ağacı sınıflandırıcısının düzenlenmesinde kullanılan iki yöntem vardır (Breiman vd., 1984; Mingers, 1989). Bunlardan ilki eğitim setinin daha ileriye bölünüp bölünmemesine karar verilen ön budama, diğeri ağaç yapısının budama yapılmaksızın oluşturulduğu daha sonra budamaların yapıldığı son budama yöntemleridir. C4.5 algoritmasının kullandığı ön budama yöntemi daha az hesaplama içermesi, veri setinin ayrılması için en iyi yolu araştırması, bilgi kazancı ya da hata azalımı gibi faktörler ile bir noktadan daha ileriye ayırım yapıp yapılmayacağını değerlendirilmesi gibi önemli avantajlara sahiptir. Bu değerlendirme belirli bir eşik değerinin altına düştüğünde bölünme kabul edilmez ve veri için en uygun yaprak olduğuna karar verilir (Breiman vd., 1984).

#### 4.3. Sınıflandırıcıların Birleştirilmesi ve İyileştirilmiş Karar Ağaçları

Son yıllarda uzaktan algılama literatüründe birçok makalede birden çok sınıflandırıcının birleştirilmesi ile tek bir sınıflandırıcının elde edilmesi düşüncesi vurgulanmaktadır (Dietterich, 2000; Zhou ve Tang, 2003; Altıncay, 2007). Sınıflandırıcıların birleştirilmesi, yeniden örneklenen eğitim setleri ile sınıflandırıcıların ayrı ayrı eğitilmesi ve sonuçta ortaya çıkan tahminler ile sınıflandırma işleminin gerçekleştirilmesi işlemlerini içerir. Birleştirme sonucu elde edilen sınıflandırıcı ile yapılan sınıflandırma doğruluğunun genel olarak her bir sınıflandırıcının tekil olarak kullanılmasından daha iyi olduğu ifade edilmektedir (Opitz ve Maclin, 1999; Pal ve Mather, 2003). Hızlandırma ve torbalama yöntemleri birleştirme işleminde kullanılan en popüler yöntemlerdir (Freund ve Schapire, 1996; Schapire, 1999; Breiman, 1996). Her iki yöntemde birleştirme işleminde kullanılacak her bir sınıflandırıcı için farklı eğitim setlerinin oluşturulmasında yeniden örnekleme tekniklerini kullanmaktadır. Bu iki yöntemin kullanımı arasındaki en temel fark hızlandırma yönteminin sürekli olarak tekrarlı bir şekilde birden çok sınıflandırıcı oluşturmasıdır. Birleştirme işleminde kullanılan diğer bir yöntem çoklu sınıflandırma yöntemidir. Çoklu hızlandırma yöntemi Adaboost yöntemi ile torbalama yönteminin birleşimi olarak düşünülebilir (Webb, 2000).

Hızlandırma yönteminin esası bir dizi sınıflandırıcı serisi oluşturulmasına dayanmaktadır (Freund ve Schapire, 1996; Nanni ve Lumini, 2006). Serinin üyesi olan her bir sınıflandırıcı için kullanılan eğitim seti, serideki bir önceki sınıflandırıcı veya sınıflandırıcıların performansına göre belirlenir. Literatürde hızlandırma algoritması ile ilgili çeşitli yöntemler kullanılmaktadır. Bu algoritmalarından en çok kullanılanı Adaboost algoritmasıdır (Quinlan, 1996; Maclin ve Opitz, 1997, Dietterrich, 2000).

Torbalama yöntemi, birleştirme için eğitim setinin rastgele bir şekilde yeniden oluşturulması ile ayrı ayrı her bir sınıflandırıcının eğitimini gerçekleştiren ön yüklemeli bir birleştirme metodudur (Breiman, 1996; Efron ve Tibshirani, 1993). Torbalama yönteminde, hızlandırma yönteminde olduğu gibi eğitim setinin yeniden örneklenmesinde bir önceki sınıflandırıcının performansı dikkate alınmamaktadır. Bu yöntemde her bir sınıflandırıcının kullandığı eğitim seti rastgele örnekleme ile oluşturulur. Orijinal eğitim setinin boyutunun  $N$  olduğu düşünülürse yeni oluşturulacak eğitim setinde  $N$  boyutlu olacaktır. Buna karşın yeni eğitim seti içerisinde orijinal veri setindeki birçok örnek tekrar edilirken, orijinal veri setindeki bazı örnekler ise hiç bulunmayacaktır. Rastgele örnekleme yöntemiyle oluşturulan farklı eğitim verileri ile birleştirme işlemindeki bir sınıflandırıcının eğitimi tamamlanır (Breiman, 1996).

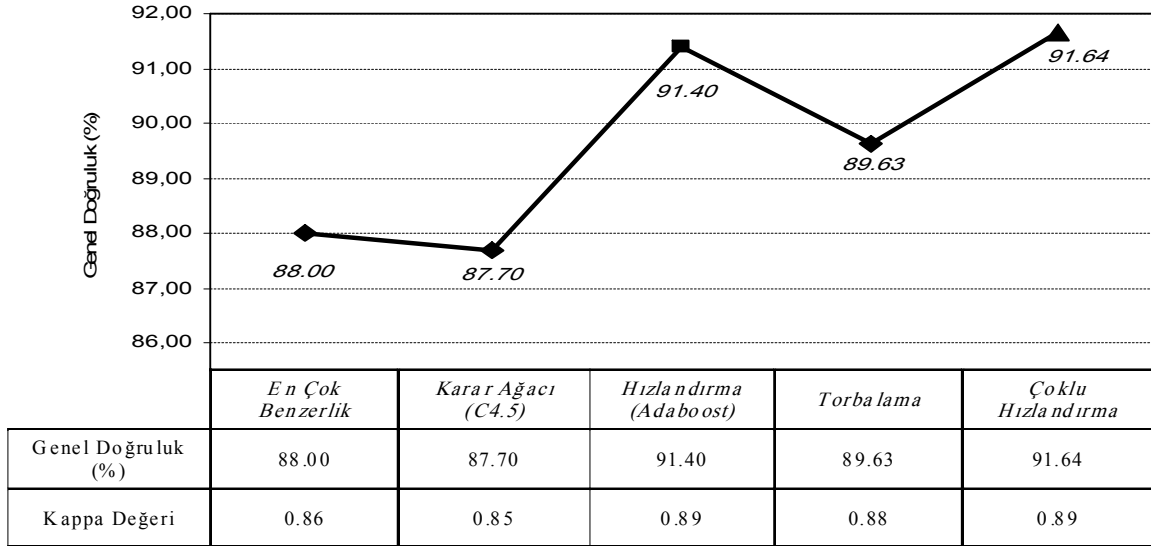
## 5. UYDU GÖRÜNTÜSÜNÜN SINIFLANDIRILMASI VE ANALİZİ

Görüntü sınıflandırma, görüntüyü oluşturan her bir pikselin tüm bantlardaki değerlerinin diğer pikseller ile karşılaştırılarak benzer piksellerin kullanıcının belirlediği bilgi kategorileri dahilinde sınıflara ayrılması işlemidir (Campbell, 1996). Objelerin, sınıf olarak adlandırılan kategorilerden birine atanmasına örüntü sınıflandırması denir. Özel bir örüntü sınıflandırma probleminde bir sınıflandırıcı sınıflandırılmak istenilen nesnelere makul düzeyde bir doğrulukla sınıflandıracak şekilde oluşturulur. Girdi ve çıktı çiftleri kullanılarak yapılan eğitim yardımıyla sınıflandırıcılar girdi verilerini verilen sınıflardan birine atamak için karar fonksiyonlarını oluşturur. Günümüze kadar uydu görüntülerinin sınıflandırılmasında birçok yöntem geliştirilmiş ve kullanılmıştır.

Çalışma alanını kapsayan uydu görüntüsünün sınıflandırılması işlemi için öncelikle arazideki mevcut durumu en iyi yansıtabilecek şekilde belirlenen altı sınıfa kapsayan eğitim alanları tespit edildi. Belirlenen eğitim alanlarını kapsayan görüntü katmanı yazılan bir MATLAB programı yardımıyla rastgele örnekleme prensibinden hareketle örneklendirilerek eğitim ve test veri setleri oluşturuldu. Eğitim ve test veri setleri belirlenirken tüm sınıflar için eşit sayıda örnek seçilerek özellikle test verisi için hesaplanan doğruluk değerlerinde oluşabilecek “önyargı” giderildi. Sonuç olarak sınıflandırmada kullanılmak üzere, 2002 tarihinde algılanan Terra ASTER görüntüsü için eğitim verisi olarak toplam 2.400 piksel, test verisi olarak toplam 1.350 piksel belirlenmiştir.

En çok benzerlik (EÇB) yöntemiyle sınıflandırma işlemi MATLAB yazılımında hazırlanan bir program ile gerçekleştirilmiştir. Sınıflandırma işleminde Terra ASTER görüntüleri için test veri seti kullanılarak doğruluk analizi gerçekleştirilmiştir. Sınıflandırma sırasında piksellerin belirlenen sınıflardan veya bu sınıfların dışında bir sınıftan olduğunun tespiti için eşik değeri %50 olarak belirlenmiştir. Söz konusu uydu görüntüsünün EÇB yöntemiyle sınıflandırılması sonucunda genel sınıflandırma doğruluğu Şekil 3’den de görüleceği üzere %88 olarak hesaplanmıştır.

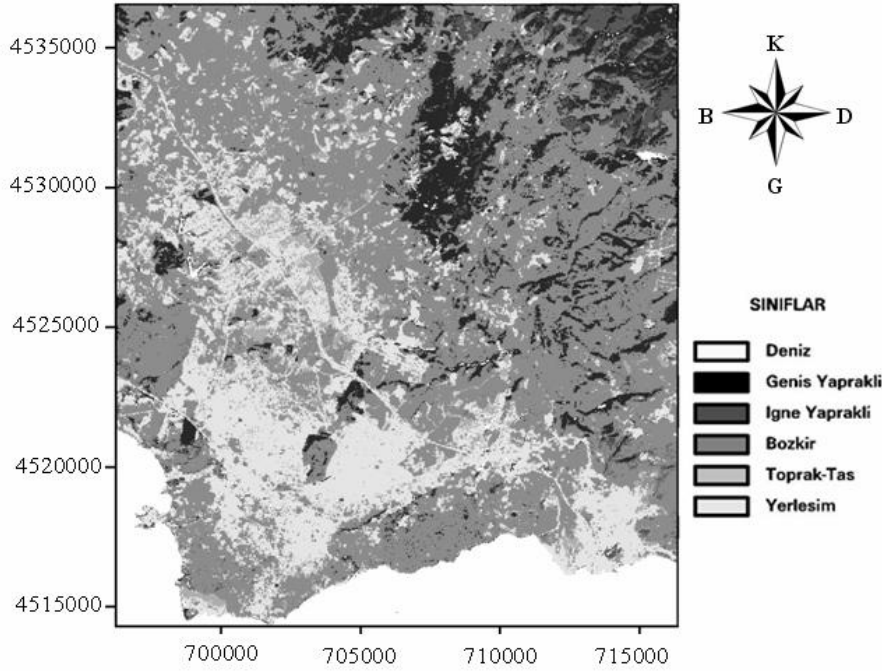
### *İyileştirilmiş Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması*



Şekil 3: Terra ASTER görüntüsünün sınıflandırılmasında kullanılan yöntemler ve elde edilen genel sınıflandırma doğrulukları

2002 tarihli uydu görüntüsünün karar ağacı sınıflandırma yöntemi ile sınıflandırılması tek değişkenli karar ağacı algoritmalarından C4.5 ile gerçekleştirilmiştir. Karar ağacının oluşumunda dallanmaya esas olacak özneliklerin seçiminde kazanç oranı kriteri hesaplamalarda kullanılmıştır. Oluşturulan karar ağacının sadeleştirilmesi ve sınıflandırma doğruluğunun artırılması amacıyla ön budama yöntemi ve budama sırasındaki hataların tahmini için karamsar budama tekniğinden yararlanılmıştır. Şekil 3'te görüleceği üzere C4.5 sınıflandırıcı ile yapılan sınıflandırma sonucunda genel sınıflandırma doğruluğu %87.7 olarak hesaplanmıştır. Bu sonuç en çok benzerlik yöntemi ile elde edilen genel doğrulukla yaklaşık olarak aynı seviyededir. Karar ağacı sınıflandırıcısının sınıflandırma performansının artırılmasına yönelik olarak sınıflandırıcıların birleştirilmesi işlemi gerçekleştirilmiştir. Sınıflandırıcıların birleştirilmesinde hızlandırma, torbalama ve bu iki yöntemin birleşimi olarak ifade edilen çoklu hızlandırma yöntemleri C4.5 algoritmasına ayrı ayrı uygulanmış ve elde edilen genel sınıflandırma doğrulukları Şekil 3'te gösterilmiştir. Hızlandırma yöntemiyle elde edilen genel sınıflandırma doğruluğu %91.40 olarak hesaplanmıştır. Uygulamada C4.5 karar ağacı algoritmasının sınıflandırma doğruluğunun artırılması amacıyla kullanılan ikinci iyileştirme yöntemi torbalama yöntemidir. Bu yöntemle elde edilen genel sınıflandırma doğruluğu %89.63 olarak hesaplanmıştır. Çalışmada son olarak kullanılan birleştirme yöntemi çoklu hızlandırmadır. Çoklu hızlandırmada temel sınıflandırıcı olarak C4.5 algoritması kullanılarak uydu görüntüsünün sınıflandırılması sonucu elde edilen genel sınıflandırma doğruluğu Şekil 3'den de görüleceği üzere %91.64 olarak hesaplanmıştır.

Şekildeki genel doğruluklar incelendiğinde hızlandırma, torbalama ve çoklu hızlandırma algoritmalarının kullanılmasıyla karar ağaçlarının sınıflandırma performansında ortalama %3'lük bir iyileştirme olduğu görülmektedir. Özellikle hızlandırma ve çoklu hızlandırma algoritmalarının birleşimi olarak ifade edilen çoklu hızlandırma algoritması ile karar ağaçlarının tekil kullanımıyla elde edilen sınıflandırma doğruluğunda yaklaşık olarak %4'lük bir iyileştirme meydana gelmiştir. Kullanılan üç iyileştirme algoritması ile hesaplanan genel doğruluklar analiz edildiğinde, hızlandırma (Adaboost) ve çoklu hızlandırma algoritmalarının yaklaşık olarak aynı (%92) sonuca sahip olduğu, bunun yanında torbalama algoritması ile elde edilen sonucun bu iki yöntemle göre yaklaşık olarak %2 düşük olduğu tespit edilmiştir. Terra ASTER uydu görüntüsünün sınıflandırılması sonucunda elde edilen tematik haritalara örnek olarak kullanılan yöntemlerden en yüksek sınıflandırma doğruluğunun hesaplandığı çoklu hızlandırma yöntemi ile elde edilen tematik harita Şekil 4' de gösterilmiştir. Bu harita incelendiğinde çalışma alanının kuzey-doğusunda özellikle yoğun orman alanlarının olduğu, şehirleşmenin ise denize yakın ve çalışma alanının orta kısmında yoğunlaştığı açıkça görülmektedir.



Şekil 4: Terra ASTER görüntüsünün çoklu hızlandırma algoritmasıyla sınıflandırılması sonucu elde edilen tematik harita

## 6. SONUÇLAR

Uzaktan algılama teknolojilerinin temel ürünü olan uydu görüntüleri üzerinden çeşitli çalışmalar için ihtiyaç duyulan yeryüzüne ait önemli bilgilerin elde edilmesinde görüntülerin sınıflandırılması söz konusudur. Sınıflandırma sonucu elde edilen tematik haritaların ve sınıflara ait bilgilerin güvenilirliği sınıflandırmada kullanılacak yöntem ve elde edilecek sınıflandırma doğruluğu ile ilişkilidir. Bu çalışmada karar ağaçlarının uydu görüntülerinin sınıflandırılmasındaki performansı popüler sınıflandırma yöntemlerinden olan en çok benzerlik sınıflandırıcısının performansı ile karşılaştırılmıştır. Diğer taraftan karar ağaçlarının sınıflandırma performansını arttırmak amacıyla hızlandırma, torbalama ve çoklu hızlandırma yöntemleri olarak bilinen üç farklı birleştirme algoritması kullanılarak uydu görüntüsü ayrı ayrı sınıflandırılarak elde edilen genel sınıflandırma doğrulukları analiz edilmiştir. Sınıflandırıcıların birleştirilmesi karar ağaçlarının sınıflandırma performansını artırılmasında kullanılan bir yöntemdir. Birleştirme işlemi kullanılan hızlandırma (Adaboost), torbalama ve çoklu hızlandırma algoritmaları ile karar ağaçlarının sınıflandırma performansında çeşitli derecelerde iyileşmeler meydana gelmektedir.

Karar ağaçları yapısı itibariyle kolay anlaşılabilir ve parametrik olmayan bir yaklaşımdır. Karar ağaçları bu özellikleri ile uzaktan algılanmış görüntülerin sınıflandırılması konusunu da içine alan farklı uygulamalarda kullanılmaktadır. Literatürde karar ağaçları ve yöntemin sınıflandırma performansını arttırmaya yönelik günümüze kadar çeşitli algoritmalar geliştirilmiş ve çeşitli derecelerde başarılar elde edilmiştir. En çok benzerlik yöntemi matematiksel ifadesinin daha basit oluşu, işlem süresinin kısalığı ve parametrik bir yaklaşım olması nedeniyle uzaktan algılanmış görüntülerin sınıflandırılmasında yaygın kullanımı olan bir sınıflandırma yöntemidir. Çalışmadan elde edilen veriler ışığında karar ağaçlarının tekil kullanımı sonucunda elde edilen sonuçların en çok benzerlik yöntemi ile benzer olduğu görülmüştür. Karar ağaçlarının tekil kullanımından ziyade birleştirme yöntemleriyle beraber kullanımı sonucunda ise yöntemin performansında seçilen birleştirme yöntemine göre değişik oranlarda iyileşmeler meydana geldiği ve en çok benzerlik yöntemine göre daha yüksek doğruluklu sonuçlar elde edildiği görülmüştür. Elde edilen sonuçlar; karar ağaçlarının özellikle çeşitli algoritmalar kullanılarak iyileştirilmiş karar ağaçlarının uzaktan algılanmış görüntülerin sınıflandırılmasında etkili ve alternatif bir yöntem olduğunu ortaya koymuştur.

## KAYNAKLAR

- Altıncay, H., 2007. *Decision trees using model ensemble-based nodes*. Pattern Recognition, sayı: 40, 3540-3551.
- Breiman L., 1996. *Bagging predictors*, Machine Learning, sayı: 24, sayfa: 123-140.
- Breiman L., Friedman J.H., Olshen R.A. and Stone C.J., 1984. *Classification and Regression Trees*, Monterey, CA: Wadsworth
- Brodley C.E., Utgoff P.E., 1995. *Multivariate Decision Trees*, Machine Learning, sayı:19, sayfa45-77.

- Campbell J.B.,1996. *Introduction to Remote Sensing*, Second Edition, The Guilford Press, New York, ABD.
- DeFries R., Hansen M., Townshend J. R. G., Sohlberg R., 1998. *Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers*, International Journal of Remote Sensing, sayı: 19, sayfa: 3141–3168.
- Dietterich T. G., 2000. *An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization*, Machine Learning, sayı: 40, sayfa: 139 - 157.
- Efron B. and Tibshirani R.J., 1993. *An Introduction to the Bootstrap*, Chapman & Hall, New York.
- Freund Y., Schapire R. E., 1996. *Experiments with a new boosting algorithm*, In: Saïtta, L. (Ed.), Proceedings of the Thirteenth International Conference on Machine Learning (ICML-96), Morgan Kaufmann, 148-156.
- Friedl M.A., Brodley C.E., 1997. *Decision tree classification of land cover from remotely sensed data*, Remote Sensing of Environment, sayı: 61, sayfa: 399–409.
- Hall F. G., Townshend J. R.G., Engman E. T., 1995. *Status of Remote Sensing Algorithms for Estimation of Land Surface State Parameters*, Remote Sensing of Environment, sayı: 51, sayfa: 138–156.
- Hansen M.C., DeFries R.S., Townshend J.R.G., Sohlberg, R., 2000. *Global land cover classification at 1 km spatial resolution using a classification tree approach*, International Journal of Remote Sensing, sayı: 21, sayfa: 1331–1364.
- Huang C., Davis L.S., Townshend, J.R.G., 2002. *An assessment of support vector machines for land cover classification*, International Journal of Remote Sensing , sayı: 23, sayfa: 725–749.
- Lu D., Weng Q., 2007. *A Survey of Image Classification Methods and Techniques for Improving Classification Performance*, International Journal of Remote Sensing, sayı: 28, sayfa: 823–870.
- Maclin R., Opitz D., 1997. *An empirical evaluation of bagging and boosting*, In: Proceedings of the Fourteenth National Conference on Artificial Intelligence. AAAI Press, Providence, RI, 546-551.
- Mather , P.M., 1987. *Computer Processing of Remote-Sensed Images*, John Wiley and Sons Ltd, 125 s.
- Mingers J., 1989. *An empirical comparison of pruning methods for decision tree induction*, Machine Learning, sayı: 4, sayfa: 227–243.
- Nanni L., Lumini A., 2006. *An experimental comparison of ensemble of classifiers for biometric data.*, Neurocomputing, sayı: 69, 1670-1673.
- Opitz D., Maclin R., 1999. *Popular ensemble methods: An empirical study*, Journal of Artificial Intelligence Research, sayı: 11, sayfa: 169–198.
- Özkan, Y., 2008. *Veri madenciliği yöntemleri*, Papatya yayıncılık eğitim, İstanbul, 216 s.
- Pal M., Mather P.M., 2003. *An assessment of the effectiveness of decision tree methods for land cover classification*, Remote Sensing of Environment, sayı: 86, sayfa: 554-565.
- Pal, M., 2007. *Ensemble learning with decision tree for remote sensing classification*, In: Proceedings of World Academy of Science, Engineering and Technology, sayı: 26, sayfa: 735–737.
- Quinlan J. R., 1987. *Simplifying decision trees*, International Journal of Man-Machine Studies, sayı: 27, sayfa: 221-234.
- Quinlan J. R., 1993. *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA.
- Quinlan J. R., 1996. *Bagging, boosting, and C4.5*, In: Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96). Portland, OR, 725-730.
- Richards J.A., 1993. *Remote Sensing Digital Image Analysis: An Introduction*, Second Edition, Springer-Verlag, Berlin.



Safavian S.R., Landgrebe D., 1991. *A Survey of Decision Tree Classifier Methodology*, IEEE Transactions on Systems Man and Cybernetics, sayı: 21, sayfa: 660-674.

Schapire R.E., 1999. *Theoretical views of boosting and applications*, In: Proceedings of the Tenth International Conference on Algorithmic Learning Theory, 13-25.

Townshend J. R. G., 1992. *Land cover*, International Journal of Remote Sensing, sayı: 13, sayfa: 1319–1328.

Webb G.I., 2000. *MultiBoosting: A technique for combining boosting and wagging*, Machine Learning, sayı: 40, sayfa: 159-196.

Zhou Z.H., Tang W., 2003. *Selective ensemble of decision trees*, Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, sayı: 2639, sayfa: 476-483.

URL 1, Gebze Belediyesi İnternet sistesi, <http://www.gebze.bel.tr>, 10 Ocak 2009.