

**Makale
(Article)**

Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği

Taşkın KAVZOĞLU*, İsmail ÇÖLKESEN*

* Gebze Yüksek Teknoloji Enstitüsü, Jeodezi ve Fotogrametri Mühendisliği Bölümü, Gebze, Kocaeli
kavzoglu@gyte.edu.tr, icolkesen@gyte.edu.tr

Özet

Uzaktan algılanmış uydu görüntüleri sağladığı önemli avantajlar sayesinde birçok çalışma için öncelikli veri kaynağı olmuştur. Uydu görüntülerinden yeryüzüne ait çeşitli bilgilerin elde edilmesinde en çok başvurulan yöntem görüntülerin sınıflandırılmasıdır. Sınıflandırma sonucu elde edilen tematik haritaların doğruluk ve güncellik derecesi sonuçlar üzerinde doğrudan etkilidir. Bu amaca yönelik olarak günümüze kadar birçok sınıflandırma algoritması geliştirilmiştir. Uydu görüntülerin sınıflandırılmasında son yıllarda kullanılmaya başlayan karar ağaçları, akış şemalarına benzeyen yapılarıyla birçok alanda başarıyla kullanılan bir kontrollü sınıflandırma yöntemidir. Yöntemin parametrik olmayan yapısı ve problem çözümündeki hızı, kullanımını yaygın hale getirmiştir. Bu çalışmada, karar ağaçlarının sınıflandırma performansı güncel bir Landsat ETM+ uydu görüntüsü kullanılarak detaylı şekilde analiz edilmiştir. Yöntemin sınıflandırma performansı yaygın kullanıma sahip en çok benzerlik yönteminin performansı ile karşılaştırılmıştır. Çalışmada kullanılan veri seti için sınıflandırma yöntemlerinin performansları istatistiksel olarak Z testi ile analiz edilmiştir. Elde edilen sonuçlar karar ağaçlarının uydu görüntülerinin sınıflandırılmasında etkin bir yöntem olduğunu göstermiştir.

Anahtar Kelimeler: Sınıflandırma, Tematik Harita, En Çok Benzerlik, Karar Ağaçları, Z test

Classification of Satellite Images Using Decision Trees: Kocaeli Case

Abstract

Remotely sensed images have been primary data sources for many studies since they provide distinctive advantages. For the extraction of various geo-information from satellite images classification is the most widely used approach. The levels of accuracy and currency of the thematic maps produced through classification have direct impacts on the results. Many algorithms have been developed in the literature for this purpose. Decision trees with flowchart-like structures that have been recently employed in the classification of satellite images are supervised classification techniques successfully used in many fields. Their non-parametric nature and process speed in problem solving have made decision trees widely used and preferred methods. In this study, classification performance of the decision trees was thoroughly analysed using a recent Landsat ETM+ imagery. Performance of decision trees was compared with the maximum likelihood classifier that has been the most widely used classifier. Performances of the classifiers for the data set considered in this study were statistically evaluated using Z test. Results show that decision trees are effective in the classification of satellite images.

Keywords: Classification, Thematic Map, Maximum Likelihood, Decision Trees, Z test

1. GİRİŞ

Uzaktan algılama teknolojileri, yeryüzünü çeşitli zaman aralıklarında farklı çözünürlüklerde görüntüleyebilmekte dolayısıyla ihtiyaç duyulan bilgilerin hızlı ve güvenilir bir biçimde elde edilebilmesine imkan sağlamaktadır. Ham halde elde edilen uydu görüntülerinden yeryüzüne ait bilgilerin

Bu makaleye atıf yapmak için

Kavzoğlu, T., Çölkesen, İ., "Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği" Harita Teknolojileri Elektronik Dergisi 2010, 2(1) 36-45

How to cite this article

Kavzoğlu, T., Çölkesen, İ., "Classification of Satellite Images Using Decision Trees: Kocaeli Case" Electronic Journal of Map Technologies, 2010, 2(1) 36-45

elde edilmesi için çeřitli istatistiksel analizler ve istatistiksel yorumlama teknikleri kullanılmaktadır. Uzaktan algılanmıř verilerin bilgiye dnüřtürölerek yeryüzüne ait bilgilerin elde edilmesinde kullanılan en önemli yöntem görüntülerin sınıflandırılmasıdır. Sınıflandırma işleminin genel amacı yeryüzü üzerinde aynı spektral özellikleri taşıyan nesnelere gruplandırılmasıdır. Diđer bir ifadeyle, görüntüyü oluşturan her bir pikselin tüm bantlardaki deđerlerinin diđer pikseller ile karşılaştırılarak benzer piksellerin kullanıcının belirlediđi sınıflara ayrılması işlemidir [1]. Sınıflandırma işleminde genel olarak kontrollü ve kontrolsüz sınıflandırma olarak adlandırılan iki yaklaşım kullanılmaktadır. Eğitim veri seti kullanılarak sınıflandırma işlemini gerçekleřtiren kontrollü sınıflandırma yaklaşımı daha yüksek doğruluklar elde edildiđinden en çok tercih edilen yöntemdir.

Günümüze kadar uzaktan algılanmıř görüntülerin sınıflandırılması amacıyla birçok sınıflandırma algoritması geliştirilmiřtir [2-3-4]. Bu algoritmalar arasında günümüzde geçerliliđi kabul edilen ve yaygın kullanıma sahip olan en çok benzerlik (EÇB) sınıflandırıcısıdır. Sınıflandırma sonucunda elde edilen tematik haritaların doğruluđu büyük veya küçük ölçeklerde gerçekleştirilen birçok alıřma için önemli bir kaynak durumundadır. Bu nedenle literatürde sınıflandırma doğruluđunun artırılması ve daha güvenilir sonuçların elde edebilmesi amacıyla gelişmiř sınıflandırma yaklaşımları ortaya atılmaktadır. Bu yaklaşımlardan başlıcaları olarak yapay sinir ađları, destek vektör makineleri ve bulanık mantık verilebilir. Bunların uzaktan algılanmıř görüntülerin sınıflandırmasındaki etkinliđi birçok alıřmada ortaya konmuřtur [5-6-7-8-9]. Bu metodların kısıtlı eğitim verisi olduđu durumlarda iyi sonuçlar üretilebildiđi de arařtırmalarla ortaya konmuřtur [10]. Son yıllarda uzaktan algılanmıř görüntülerin sınıflandırılması üzerine yapılan alıřmalarda kullanılan karar ađaçları (KA), hesaplama açısından hızlı ve istatistiksel kabullere dayanmayan bir sınıflandırma yöntemidir [5-11]. Karar ađaçları akıř řemalarına benzeyen bir yapıya sahip olup genel olarak sınıflandırma probleminin kademeli bir řekilde özölmesi prensibine dayanmaktadır.

Bu alıřmada, karar ađaçları ile 2009 tarihli Landsat ETM+ uydu görüntüsü kullanılarak Kocaeli iline ait arazi kullanımının belirlenmesi amaçlanmıřtır. Karar ađaçlarının sınıflandırılma performansı, görüntü sınıflandırmada temel ölçüt ya da dayanak olarak kabul edilen en çok benzerlik sınıflandırıcısı ile karşılaştırılmıřtır. Sınıflandırma sonucunda elde edilen genel doğrulukların ve Kappa deđerlerinin karşılaştırılmasının yanında sınıflandırıcıların performansındaki farkların istatistiksel olarak anlamlı olup olmadıđının arařtırılması amacıyla yaygın kullanıma sahip Z testi uygulanmıřtır.

2. ALIřMA ALANI

Kocaeli, Marmara denizinin doğusunda ve yaklaşık 3.505 km²'lik yüzölçümüyle Türkiye'nin en büyük sanayi kuruluşlarına ev sahipliđi yapmaktadır (řekil 1). Bir sanayi kenti olan Kocaeli başta merkez ilçesi İzmit olmak üzere 11 ilçeye sahiptir. Türkiye'nin en büyük metropoliten řehri olan İstanbul'un sınırında bulunması, çevre yolu ve diđer bađlantı yolları üzerinde olması, Karadeniz ve Marmara denizine olan kıyıları ve sahip olduđu limanlar nedeniyle Kocaeli bir sanayi kenti haline dönuřmüřtür. Kocaeli ilinin 1970 yılında 385.408 olan nüfusu 2009 yılı sonu itibariyle 1.522.408 ulaşmıřtır (www.tuik.gov.tr).

3. MATERYAL VE YÖNTEM

Bu alıřmada, arazi örtüsünün sınıflandırılması için alıřma alanını kapsayan 30.09.2009 tarihinde kaydedilmiř Landsat ETM+ görüntüsü kullanılmıřtır. Görüntünün geometrik olarak düzeltilmesi ve UTM koordinat sistemine dönuřümünde bölgeye ait 1/25.000 ölçekli haritalar ve mevcut halihazır haritalar kullanılmıřtır. Uydu görüntülerinin yeniden örneklenmesinde orijinal deđerleri koruduđundan en yakın komřu yöntemi tercih edilmiřtir. Ayrıca, görüntünün sınıflandırılmasında esas teşkil edecek eğitim alanları farklı tarihlerde çekilmiř hava fotođrafları, meřcere haritaları ve arazide el GPS aleti ile belirlenen noktalar yardımıyla belirlenmiřtir. Yapılan ön alıřmalar neticesinde alıřma alanında altı temel sınıfın (su, geniş yapraklı orman, iđne yapraklı orman, bozkır, toprak-tař ve yerleşim) mevcut

olduğu belirlenmiş ve bu altı sınıf için sınıflandırmada kullanılmak üzere eğitim ve test verileri oluşturulmuştur.



Şekil 1: Çalışma alanı olarak seçilen Kocaeli il sınırları

Çalışmada kullanılan yöntemlerin sınıflandırma performanslarının değerlendirilmesinde öncelikli olarak sınıflandırma sonucunda elde edilen genel doğruluklar ve Kappa değerleri kullanılmıştır. Bu değerlendirmelerin yanında sınıflandırıcıların (Karar ağaçlar ve EÇB) performansları arasındaki farkların istatistiksel olarak anlamlılığının analizi için Z testi kullanılmıştır. Z testi sınıflandırma sonucu elde edilen Kappa değerlerinin karşılaştırılması amacıyla kullanılmaktadır. Uydu görüntüsünün sınıflandırılmasıyla elde edilen iki tematik haritanın doğruluğu arasındaki farkın istatistiksel olarak değerlendirilmesi genellikle her bir tematik harita için hesaplanan Kappa değerlerinin karşılaştırılmasına dayanmaktadır [12]. Söz konusu iki bağımsız Kappa değeri arasındaki farkın istatistiksel olarak anlamlılığı aşağıdaki şekilde hesaplanan Z değeri ile test edilir.

$$Z = \frac{K_1 - K_2}{\sqrt{\sigma(K_1) + \sigma(K_2)}} \quad (1)$$

Bu eşitlikte K_1 ve K_2 karşılaştırılan iki Kappa değerini, $\sigma(K_1)$ ve $\sigma(K_2)$ ise Kappa değerleri için hesaplanan varyansları göstermektedir. Z test değeri normal dağılımdadır. Hesaplanan Z değeri önceden belirlenen güven aralığındaki ($Z_{\alpha/2}$) kritik değer ile karşılaştırılır. Hesaplanan Z değeri %95 güven aralığındaki $Z=1.96$ kritik değerinden büyük olduğunda iki sınıflandırma performansı arasındaki farkın istatistiksel olarak anlamlı olduğu söylenebilir. Bu durum iki bağımsız Kappa değerinin istatistiksel olarak farklı olduğunu, dolayısıyla iki sınıflandırıcıya ait performansların birbirinden farklı olduğunu gösterir.

4. EN ÇOK BENZERLİK SINIFLANDIRMA YÖNTEMİ

En çok benzerlik yöntemi literatürde en yaygın olarak kullanılan bir sınıflandırma yöntemidir. Bu yöntemde ortalama değer, varyans ve kovaryans gibi istatistiki değerlerin tümü dikkate alındığından parametrik bir yaklaşımdır. 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 yöntemde, sınıf kontrol verilerini oluşturan noktalar kümesindeki dağılımın, normal

dağılımda olduğu kabulü yapılır. 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 matrisini de dikkate alarak sınıflandırır. Bu şekilde örnek piksellerin özellik uzayındaki dağılımları da dikkate alınmış olur [13].

Pikselin p sayıdaki bant değerlerini içeren x vektörünün bir k sınıfında olma olasılığı aşağıdaki şekilde hesaplanır.

$$P(x)=2\pi^{-0.5p} |S_i|^{-0.5} \exp\left[-0.5\left(y^T S_i^{-1} y\right)\right] \quad (2)$$

Burada $P(x)$ olasılık değerini, S_i i . sınıf için varyans-kovaryans matrisini, $||$ matrisin determinantını, $y=(x-\bar{x}_i)$ ve \bar{x}_i ise i . sınıf için ortalama değeri ifade eder.

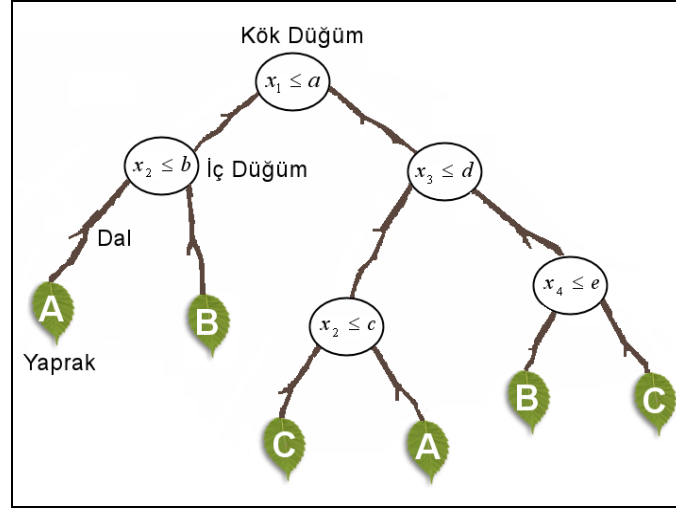
5. KARAR AĞAÇLARI SINIFLANDIRMA YÖNTEMİ

Karar ağaçları (KA) son yıllarda literatürde yaygın kullanımı olan bir sınıflandırma ve örüntü tanımlama algoritmasıdır. Bu yöntemin yaygın olarak kullanımının en önemli nedeni ağaç yapılarının oluşturulmasında kullanılan kuralların anlaşılabilir ve sade olmasıdır. KA sınıflandırma işleminin gerçekleştirilmesinde çok aşamalı veya ardışık bir yaklaşım kullanılmaktadır. Yöntem uydu görüntülerinin sınıflandırılmasında olduğu gibi çok karışık bir sınıflandırma problemini aşamalı bir hale getirerek basit bir karar verme işlemi gerçekleştirir [14].

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 öznitelik (uydu görüntüsü için bant değeri) 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 [15]. Başka bir ifadeyle bir ağaç yapısı; verileri içeren bir kök düğümü, iç düğümler (dallar) ve uç düğümlerden (yapraklar) oluşur. Eğitim verilerine ait öznitelik 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. Ağacın ilk düğümü olan kök düğümünde verilerin sınıflandırılması ve ağaç yapısının oluşturulması için sorular sorulmaya başlanır ve dalları olmayan düğümler ya da yapraklar bulunana kadar bu işlem devam eder [11]. Oluşturulan ağacın 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. Kökten her bir yaprağına giden tek bir yol veya tek bir karar kuralı vardır. Şekil 2’de üç sınıfa ait dört boyutlu öznitelik değerlerinden oluşan basit bir ağaç yapısı görülmektedir. Şekilde x_i öznitelik değerlerini (uydu görüntüsü için bant değerlerini); a, b, c, d ve e değerleri dallanma için esas olan eşik değerleri ve A, B ve C ise sınıf etiketlerini göstermektedir.

Ağaç oluşumunun her bir aşamasında kullanılan değişkenlerin sayısına bağlı olarak tek değişkenli veya çok değişkenli karar ağaç yapıları mevcuttur [16]. Tek değişkenli karar ağaçları genellikle arazi örtüsü sınıflandırmalarında kullanılan ağaç yapılarıdır [11-17]. Tek değişkenli karar ağacının her bir iç düğümünde tek bir özneliğin test edilmesi ile verinin iki veya daha fazla alt kümeye ayrılması söz konusudur [18]. Bu şekilde karar ağacı tekrarlı bir şekilde girdi verisine ait bir yaprak düğümüne ulaşıncaya kadar ve yapraktaki sınıf etiketi gözlemlere atanıncaya kadar bölünme devam eder. Tek değişkenli bir karar ağacında karar sınırlarının spesifik değerleri eğitim verisinden deneysel olarak tahmin edilir. Sürekli verilerin olması durumunda, eğitim verisinden her bir iç düğüm noktasında x_i 'nin veri uzayındaki bir

özelliği gösterdiği ve c 'nin x_i 'nin gözlenen aralığında bir eşik değeri olduğu $x_i > c$ şeklindeki mantıksal sınama gerçekleştirilir. Farklılıkların maksimum hale getirilmesi veya iniş düğümlerindeki benzerliğin minimum hale getirilmesi gibi bazı koşullar kullanılarak eşik değeri c belirlenebilir.



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

Karar ağaçlarının oluşturulmasındaki en önemli adım ağaçtaki dallanmanın hangi kriterle veya kıstasa göre yapılacağı ya da hangi öznelik değerlerine göre ağaç yapısının oluşturulacağıdır. Literatürde bu problemin çözümü için geliştirilmiş çeşitli yaklaşımlar vardır. Bunlardan en önemlileri bilgi kazancı ve bilgi kazanç oranı [15], Gini indeksi [19], Towing kuralı [19] ve Ki -Kare olasılık tablo istatistiği [20] yaklaşımlarıdır. Bilgi kazancı ve bilgi kazanç oranının kullanımı [15] tarafından ortaya atılmıştır. Bu yöntemle göre karar ağacında hangi özelliğe göre dallanmanın yapılacağını belirlemek üzere entropi kurallarını içeren bilgi teorisi kullanılmıştır. Entropi bir sistemdeki düzensizliğin ya da belirsizliğin ölçüsüdür. Tek değişkenli karar ağaçlarında ID3 algoritması bilgi kazancı yaklaşımını kullanmaktadır. Bu algoritmanın geliştirilmiş hali olan C4.5 algoritması bölünme bilgisi kavramı ile bilgi kazancıdan yararlanarak hesaplanan kazanç oranı yaklaşımını kullanmaktadır [15]. Bir veri setinin C_1, C_2, \dots, C_n şeklinde birkaç sınıftan oluştuğu ve T 'nin sınıf değerlerini gösterdiği düşünülürse, bir sınıfa ait olasılık $P_i = (C_i/|T|)$ olur ve sınıflara ait entropi;

$$Entropi(T) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (3)$$

şeklinde hesaplanır. Veri setindeki B özneliğine göre T sınıf değerleri T_1, T_2, \dots, T_n şeklinde alt kümelere ayrıldığı göz önüne alınsın. B öznelik değerleri kullanılarak T sınıf değerlerinin bölünmesi sonucunda elde edilecek kazanç,

$$Kazanç(B, T) = Entropi(T) - \sum_{i=1}^n \frac{|T_i|}{|T|} Entropi(T_i) \quad (4)$$

eşitliği ile hesaplanır. T kümesi için B özneliğinin değerini belirlemede bölünme bilgisi kullanılır (Eşitlik 5).

$$Bölünme Bilgisi (B) = - \sum_{i=1}^k \frac{|T_i|}{|T|} \log_2 \left(\frac{|T_i|}{|T|} \right) \quad (5)$$

Bu durumda kazanç oranı,

$$Kazanç\ Oranı = \frac{Kazanç(B,T)}{Bölünme\ Bilgisi(B)} \quad (6)$$

şeklinde hesaplanır. Bu eşitlik, sınıflandırma işleminde kullanılacak ayırma ile elde edilen bilgi oranını verir. Bu ölçüt kullanılarak, ağacın her bir düğümünde kazanç oranı maksimum olacak şekilde T eğitim kümesi tekrarlı bir şekilde ayrılır. İşleme her bir yaprak düğümü sadece bir sınıfa ait gözlem değerleri içerene kadar tekrar edilir.

Karar ağacı oluşumunda diğer önemli bir husus oluşturulan ağaç yapısının budanmasıdır. Karar ağacı sınıflandırıcısı eğitim verisini sadece tek bir sınıf içeren alt kümelerle bölmekte, dolayısıyla çok geniş ve karmaşık bir ağaç yapısı ortaya çıkmaktadır. Bu nedenle karar ağacında, bir alt ağacı atarak yerine bir yaprak yerleştirmek söz konusu olabilir. Bu şekilde gerçekleştirilen işleme karar ağacının budanması adı verilir [21-22]. Budama ile karar ağacının sınıflandırma doğruluğunu etkilemeyen veya katkısı olmayan kısımları çıkarılır. Böylece daha az karmaşık ve daha anlaşılabilir bir ağaç elde edilir. Budama ile ağaç yapısının sadeleştirilmesi ve işlem karmaşıklığının azaltılmasında genellikle iki yöntem kullanılır [19]. Bunlardan ilki ağaç yapısı oluşturulurken küçültülmesine karar verilen ön budama yöntemi, diğeri ise ağaç yapısı oluşturulduktan sonra budamaların yapıldığı son budama yöntemidir. C4.5 karar ağaçlarında kullanılan ön budama yöntemi, daha az hesaplama içermesi, veri setinin ayrılması için en iyi yolu araştırması ve bilgi kazancını değerlendirilmesi yönlerinden önemli avantajlara sahiptir. Bu değerlendirme belirli bir eşik değerinin altına düştüğünde bölünme kabul edilmez ve veri için en uygun yaprak olduğuna karar verilir [11-19]. Eğer tek bir yaprağı olan alt ağacın veya bu ağacın en çok kullanılan dalının budanması beklenen hata oranını düşürecekse ağaç budanır. Alt dallardaki hata oranı azaldığından tüm ağaç için hata oranı azalacaktır. Budama işlemi sonunda hata oranının minimum hale getirildiği bir ağaç elde edilir.

6. UYGULAMA

Çalışma alanını kapsayan Landsat ETM+ uydu görüntüsünü sınıflandırmak için yazılan bir MATLAB programı yardımıyla rastgele örnekleme prensibinden hareketle eğitim ve test verisi hazırlanmıştır. Eğitim ve test veri setleri belirlenirken tüm sınıflar için eşit sayıda örnek seçilerek özellikle test verisi için hesaplanan doğruluk değerlerinde oluşabilecek önyargının giderilmesi ve sınıflandırma sonuçlarının karşılaştırılabilir olması amaçlanmıştır. Sonuç olarak, sınıflandırmada kullanılmak üzere eğitim verisi olarak toplam 12.000 piksel, test verisi olarak ise toplam 6.000 piksel seçilmiştir. En çok benzerlik (EÇB) ve karar ağaçları (KA) ile görüntünün sınıflandırılması işlemleri MATLAB ortamında yazılmış programlarla gerçekleştirilmiştir.

EÇB yöntemi kullanılarak yapılan sınıflandırma sonucunda elde edilen hata matrisi Tablo 1'de gösterilmiştir. Tablodan da görüleceği üzere elde edilen genel sınıflandırma doğruluğu %91,30 ve Kappa 0,89 olarak hesaplanmıştır.

Tablo 1: En çok benzerlik sınıflandırıcısı için hata matrisi

Sınıf	1	2	3	4	5	6	Üretici Doğ. (%)	Kullanıcı Doğ. (%)
1	993	2	0	0	5	0	100	99,30
2	0	975	25	0	0	0	97,99	97,50
3	0	18	918	34	29	1	92,35	91,80
4	0	0	43	926	29	2	88,19	92,60
5	0	0	7	84	746	163	84,58	74,60
6	0	0	1	6	73	920	84,71	92,00
Genel Doğruluk: %91,30							Kappa: 0,89	

1: Su, 2: Geniş yapraklı orman, 3: İğne yapraklı orman, 4: Bozkır, 5: Toprak-taş, 6: Yerleşim

Karar ağaçları ile sınıflandırma işlemi için tek değişkenli karar ağacı algoritmalarından C4.5 algoritması kullanılmıştır. Karar ağacı modeli oluşturulmasında; dallanmaya esas olacak özneliklerin seçiminde kazanç oranı kriteri, oluşturulan karar ağacının sadeleştirilmesi ve sınıflandırma doğruluğunun artırılması amacıyla ön budama yöntemi kullanılmıştır. Tespit edilen düzenleme parametre değerleri kullanılarak KA'nın eğitimi tamamlanarak test veri seti üzerinde sınıflandırma doğruluk analizi hata matrisi üzerinden yapılmıştır (Tablo 2). C4.5 sınıflandırıcı ile yapılan sınıflandırma sonucunda genel sınıflandırma doğruluğu %92,35 ve Kappa 0,90 olarak hesaplanmıştır.

Tablo 2: Karar ağacı sınıflandırıcısı için hata matrisi

Sınıf	1	2	3	4	5	6	Üretici Doğ. (%)	Kullanıcı Doğ. (%)
1	997	1	2	0	0	0	100	99,70
2	0	982	18	0	0	0	98,00	98,20
3	0	18	957	22	2	1	95,60	95,70
4	0	1	14	931	38	16	91,00	93,10
5	0	0	7	50	804	139	84,54	80,40
6	0	0	3	20	107	870	84,80	87,00

Genel Doğruluk: %92,35 Kappa: 0,90

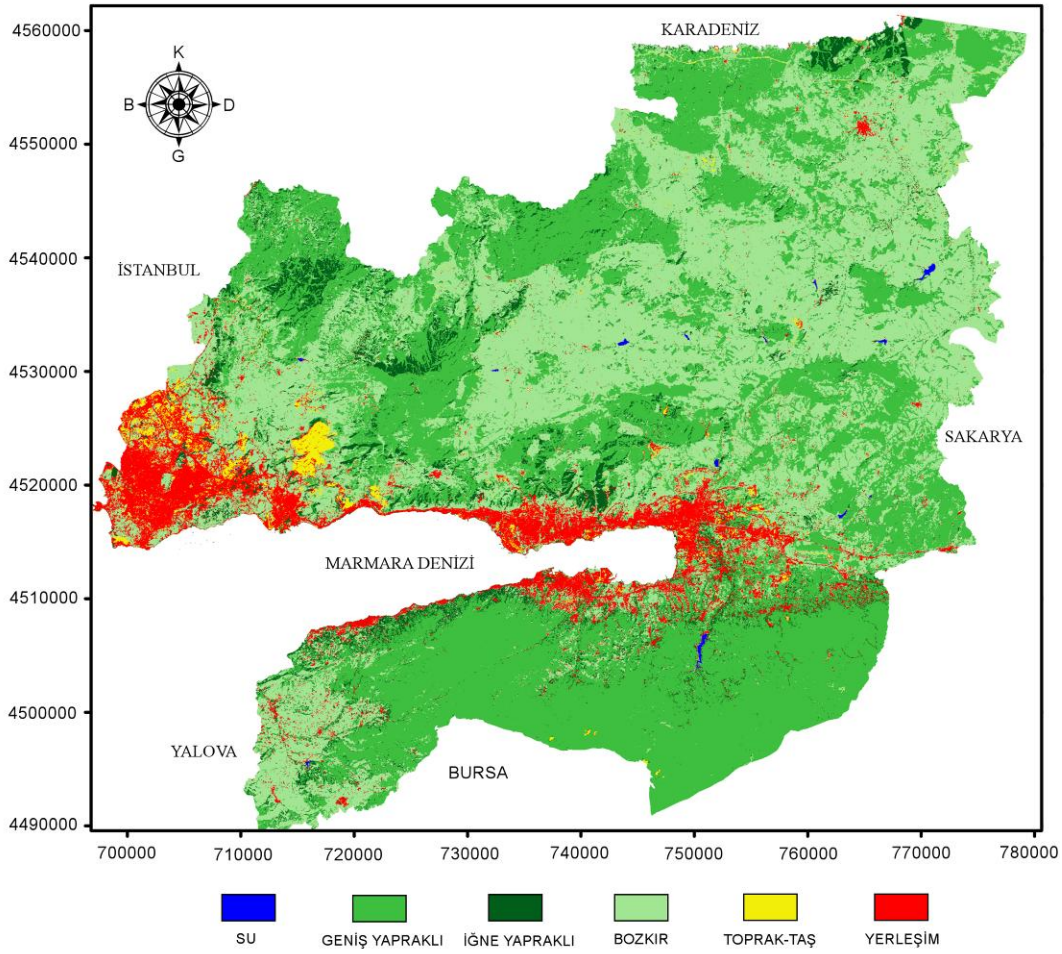
1: Su, 2: Geniş yapraklı orman, 3: İğne yapraklı orman, 4: Bozkır, 5: Toprak-taş, 6: Yerleşim

Her iki yöntem için elde edilen hata matrisleri detaylı olarak analiz edildiğinde, EÇB ve KA yaklaşımlarının toprak-taş örnek piksellerini genel olarak yerleşim ve bozkır pikselleriyle karıştırdığı görülmektedir. Spektral özellikleri benzer toprak-taş ve yerleşim piksellerinin ayırımında sınıflandırma yaklaşımlarının etkinliği sonuçlara direkt olarak yansımış durumdadır. Bu çalışmada ele alınan 30 metre konumsal çözünürlüklü görüntü düşünüldüğünde, çalışma alanındaki birçok taş-toprak ve yerleşim pikselinin "karışık" pikseller olduğu ifade edilebilir. Karar ağacı sınıflandırıcısı yerleşim sınıfı hariç diğer tüm sınıflar için daha yüksek sınıflandırma doğruluğu ürettiği görülmüştür.

Sınıflandırıcıların genel doğrulukları karşılaştırıldığında karar ağacı (KA) yöntemiyle elde edilen sınıflandırma doğruluğunun en çok benzerlik (EÇB) yöntemine göre yaklaşık olarak %1 daha yüksek olduğu görülmüştür. Bu noktada Z testi ile sınıflandırıcıların performanslarındaki bu farkın istatistiksel olarak anlamlılığı analiz edilmiştir. Eğitim veri setiyle elde edilen KA ve EÇB modellerinin test veri setine uygulanması sonucunda elde edilen Kappa değerleri ve bu değerlere ait hesaplanan varyanslar kullanılarak Z testinin uygulanması sonucunda 2.10 değeri hesaplanmıştır. Elde edilen bu istatistik değeri 0,05 güven aralığındaki $Z=1.96$ kritik değerinden büyük olduğundan iki sınıflandırıcının genel doğrulukları arasındaki %1'lik farkın istatistiksel olarak anlamlı olduğu bulunmuştur. Bu durum çalışmada kullanılan veri seti için karar ağacı yönteminin daha doğru sınıflandırma sonuçları ürettiği ve sınıflandırma performanslarının istatistiksel olarak anlamlı derecede farklı olduğu görülmüştür.

EÇB ve KA sınıflandırıcıları için oluşturulan modeller ile çalışma alanına ait Landsat ETM+ görüntüsü sınıflandırılarak Şekil 3 ve 4'te gösterilen tematik haritalar üretilmiştir. Tematik haritalardan da görüleceği üzere şehirleşmenin genellikle sahil kesimlerinde ve ana yol arterleri çevresinde gerçekleştiği görülmektedir. EÇB sınıflandırmasında toprak-taş türüne ait piksellerle yerleşim alanlarına ait piksellerin karıştığı açık şekilde görülmüştür. Bunun başlıca nedeninin yerleşim alanları çevresindeki toprak-taş örtülü alanlar olduğu düşünülmektedir. Kocaeli il sınırları içerisindeki geniş yapraklı ormanlık alanların ağırlıklı olarak şehrin güney kısmında ve iğne yapraklı ormanların ise özellikle şehrin kuzeyindeki Karadeniz kıyılarında yoğunlaştığı görülmüştür. Söz konusu tematik haritalardan karar ağacı sınıflandırıcısı ile oluşturulan tematik haritada Kocaeli ili için belirlenen arazi örtüsü sınıfları görsel olarak kolaylıkla ayırt edilebilirken EÇB sınıflandırıcısı ile oluşturulan tematik haritanın daha heterojen (karışık) görünümde olduğu belirlenmiştir. Tematik haritalar analiz edildiğinde iğne yapraklı orman ve

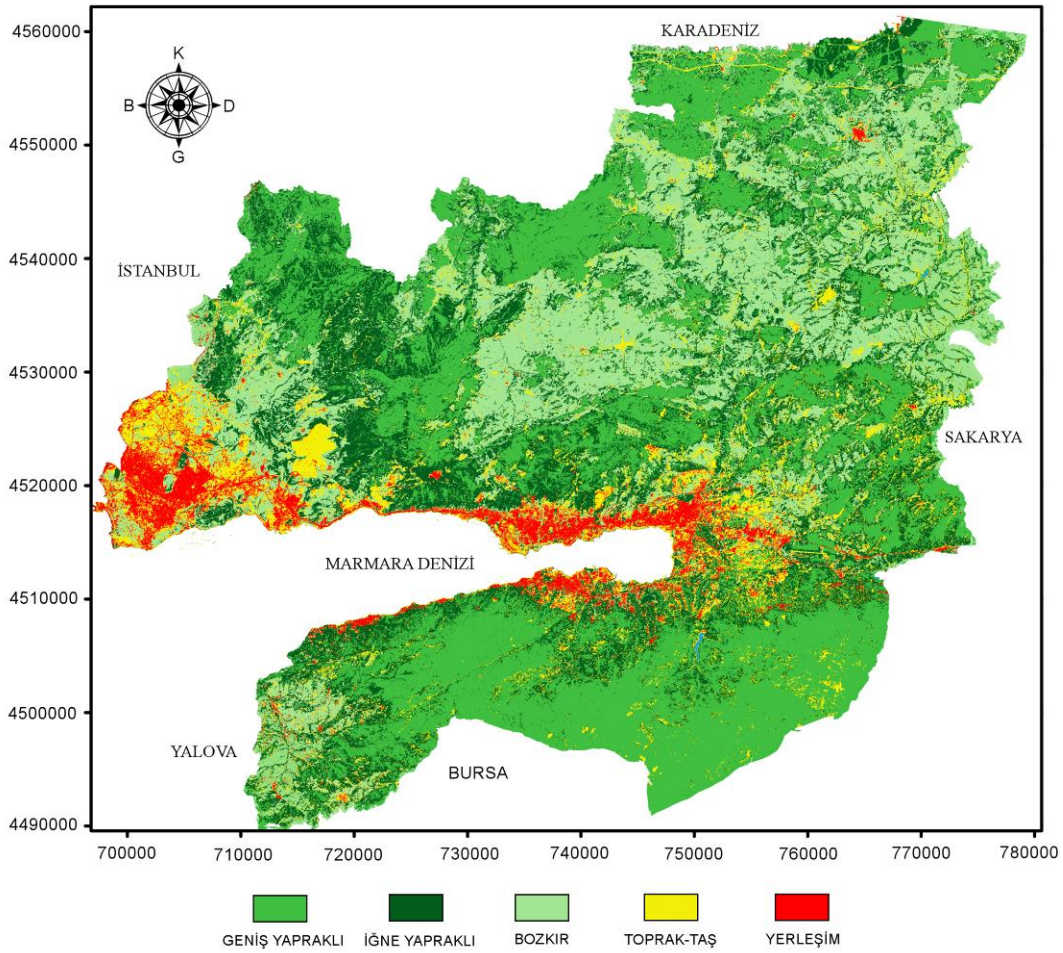
geniř yapraklı orman gibi benzer spektral özelliklere sahip piksellerin sınıflandırılmasında KA sınıflandırıcısının EB yöntemine göre daha iyi sonuçlar ürettiđi görölmüřtür.



řekil 3: Landsat ETM+ görüntüsünün karar ağacı yöntemi ile sınıflandırılması sonucu elde edilen tematik harita

7. SONU VE ÖNERİLER

Uzaktan algılanmış görüntülerin sınıflandırılması sonucu elde edilen tematik haritaların güvenilir bir kaynak olarak kullanılabilmesi için en önemli kořullardan birisi söz konusu haritaların doğruluđudur. Arazi kullanımı veya arazi örtüsünü gösteren tematik haritaların doğruluđu büyük ölçüde sınıflandırmaya esas olacak veri setinin kalitesi ve kullanılacak sınıflandırma yönteminin performansı ile ilişkilidir. Bu durum uzaktan algılama alanında çalışmalar yapan arařtırmacıları tematik harita üretiminde daha güvenilir sonuçlar üretebilecek yeni sınıflandırma yöntemleri arayışına yöneltmiştir. En çok benzerlik (EB) ve en yakın komşuluk gibi istatistiksel kabullere dayalı sınıflandırma algoritmaları literatürde yaygın olarak kullanılmalarına kořun sınıflandırma problemlerinde yetersiz kalabilmektedir. Bu amaçla son yıllarda yapay sinir ağları, destek vektör makineleri ve bulanık mantık gibi parametrik olmayan ileri sınıflandırma algoritmalarının kullanımı söz konusu olmuřtur. Karar ağaçları kolay anlaşılabilen ve hızlı çözümlene kabiliyetine sahip bir sınıflandırma yöntemi olarak son dönemde uzaktan algılama alanında kullanılan güçlü bir sınıflandırma yöntemidir.



Şekil 4: Landsat ETM+ görüntüsünün en çok benzerlik yöntemi ile sınıflandırılması sonucu elde edilen tematik harita

Bu çalışmada karar ağaęları, çalışma alanını kapsayan 2009 tarihli Landsat ETM+ görüntüsünün sınıflandırılmasında kullanılmış ve Kocaeli iline ait genel arazi örtüsünü gösteren tematik harita üretilmiştir. Uzaktan algılanmış görüntüler yardımıyla tematik harita üretiminde karar ağaęlarının sınıflandırma performansı EÇB yönteminin performansı ile karşılaştırıldığında daha etkili olduğu görülmüştür. Literatürde yaygın olarak kullanılan Z testi kullanılarak iki yöntemin sınıflandırma performansları analiz edildiğinde sınıflandırma doğrulukları arasındaki farkın istatistiksel olarak anlamlı olduğu dolayısıyla sınıflandırma performanslarının istatistiksel olarak birbirinden farklı olduğu görülmüştür. Bu bulgu ışığında karar ağaęı sınıflandırıcısının bu çalışmada kullanılan veri seti için EÇB yönteminden daha doğru sonuçlar ürettiği söylenebilir. Sonuç olarak; yapılan bu çalışma, hızlı ve parametrik olmayan bir sınıflandırıcı olan karar ağaęlarının uydu görüntülerinin sınıflandırılmasında etkin ve alternatif bir yöntem olduğunu göstermektedir.

8. KAYNAKLAR

1. Campbell, J.B., 1996, "Introduction to Remote Sensing", Guilford Press, New York, 621 s.
2. Townshend J.R.G., 1992, "Land cover", International Journal of Remote Sensing, 13, 1319–1328
3. 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, 51, 138–156
4. Lu D., Weng Q., 2007, "A survey of image classification methods and techniques for improving classification performance", International Journal of Remote Sensing, 28, 823–870

5. 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*, 23, 725–749
6. Erbek, F.S., Özkan, C., Taberner, M., 2003, "Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities", *International Journal of Remote Sensing*, 25, 1733-1748
7. Pal, M., Mather, P.M., 2005, "Support vector machines for classification in remote sensing", *International Journal of Remote Sensing*, 26, 1007-1011
8. Kavzoglul, T., Colkesen, I., 2009, "A kernel functions analysis for support vector machines for land cover classification", *International Journal of Applied Earth Observation and Geoinformation*, 11, 352-359
9. Paola, J.D., 1994, "Neural network classification of multispectral imagery", MSc Thesis, The University of Arizona, USA
10. Foody, G.M., 1995, "Using prior knowledge in artificial neural network classification with a minimal training set", *International Journal of Remote Sensing*, 16, 301-312
11. Pal M., Mather P.M., 2003, "An assessment of the effectiveness of decision tree methods for land cover classification", *Remote Sensing of Environment*, 86, 554-565
12. Foody, G.M., 2004, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy", *Photogrammetric Engineering and Remote Sensing*, 70, 627-633
13. Mather, P.M., 1987, "Computer Processing of Remote-Sensed Images", John Wiley and Sons, 125 s
14. Safavian S.R., Landgrebe D., 1991, "A survey of decision tree classifier methodology", *IEEE Transactions on Systems Man and Cybernetics*, 21, 660-674
15. Quinlan J.R., 1993, "C4.5: Programs for Machine Learning", Morgan Kaufmann, San Mateo, CA, 302 s
16. Friedl M.A., Brodley C.E., 1997, "Decision tree classification of land cover from remotely sensed data", *Remote Sensing of Environment*, 61, 399–409
17. 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*, 19, 3141–3168
18. Swain P.H., Hauska H., 1977, "Decision tree classifier - design and potential", *IEEE Transactions on Geoscience and Remote Sensing*, 15, 142-147
19. Breiman L., Friedman J.H., Olshen R.A. and Stone C.J., 1984, "Classification and Regression Trees" Monterey, CA: Wadsworth, 358 s
20. Mingers J., 1989, "An empirical comparison of pruning methods for decision tree induction", *Machine Learning*, 4, 227–243
21. Quinlan J.R., 1987, "Simplifying decision trees", *International Journal of Man-Machine Studies*, 27, 221-234
22. Özkan, Y., 2008, "Veri Madenciliđi Yöntemleri", Papatya Yayıncılık Eđitim, İstanbul, 216 s