

ÖRNEK TABANLI K-STAR ALGORİTMASI İLE UZAKTAN ALGILANMIŞ GÖRÜNTÜLERİN SINIFLANDIRILMASI

İ. Çölkesen *, T. Kavzoğlu

GYTE, Mühendislik Fakültesi, Jeodezi ve Fotogrametri Müh. Bölümü, 41400 Gebze-Kocaeli
(icolkesen@gyte.edu.tr, kavzoglu@gyte.edu.tr)

ANAHTAR SÖZCÜKLER: Arazi kullanımı, Sınıflandırma, Tematik harita, Örnek tabanlı sınıflandırma, K-Star algoritması

ÖZET:

Günümüz teknoloji ve bilişim alanındaki ilerlemelere paralel olarak kullanım alanları her geçen gün artan uzaktan algılama teknolojileri yeryüzü ile ilgili yapılan birçok çalışma için temel veri kaynağı durumundadır. Ham halde elde edilen uydu görüntüleri birtakım görüntü işleme ve yorumlama teknikleri yardımıyla değerlendirilerek yeryüzünün çeşitli özelliklerini gösteren tematik haritalar üretilmektedir. Uydu görüntüleri yardımıyla tematik haritaların üretilmesinde en sık kullanılan yöntem görüntülerin sınıflandırılmasıdır. Tematik haritaların doğruluğu sınıflandırma işleminin gerçekleştirilmesinde kullanılacak algoritmalar ile doğrudan ilişkilidir. Günümüze kadar sınıflandırma işleminin gerçekleştirilmesi amacıyla birçok yöntem geliştirilmiştir. Uzaktan algılanmış görüntülerin sınıflandırılmasında yaygın olarak kullanılan örnek tabanlı sınıflandırıcılar, pikseller arasındaki spektral uzaklıkların ya da benzerliklerin belirlenmesi ile sınıflandırma işlemini gerçekleştirmektedir. Söz konusu uzaklıkların belirlenmesinde çeşitli uzaklık (benzerlik) fonksiyonları kullanılmaktadır. Pikseller arasındaki benzerliğin belirlenmesinde entropik uzaklık fonksiyonunu esas alan K-star algoritması obje tanımlama, veri madenciliği ve kontrol sistemleri gibi birçok kullanılmakta olan örnek tabanlı bir yaklaşımdır. Bu çalışmada, K-star algoritmasının uzaktan algılanmış görüntülerin sınıflandırılmasındaki etkinliği örnek veri setleri yardımıyla detaylı olarak incelenmiştir. Uygulamada K-star algoritması ile 2009 ve 2010 tarihlerinde algılanmış Landsat TM görüntüleri kullanılarak Kocaeli ve çevresinin genel arazi kullanımının belirlenmesi amaçlanmıştır. Yöntemin sınıflandırma performansı en çok benzerlik ve en yakın komşuluk yöntemlerinin performansı ile karşılaştırılmıştır. Yöntemler için elde edilen genel sınıflandırma doğrulukları arasındaki farkların istatistiksel olarak anlamlılığı Z testi kullanılarak analiz edilmiştir. Elde edilen sonuçlar matematiksel olarak kolay ifade edilebilen K-star algoritmasının uzaktan algılanmış görüntülerin sınıflandırılmasında geleneksel yöntemlere alternatif bir yöntem olduğunu göstermiştir.

1. GİRİŞ

Dinamik bir yapıya sahip yeryüzü zaman içerisinde gerek doğal gerekse insan kaynaklı faaliyetler sonucu değişimlere uğramaktadır. Özellikle insan kaynaklı faaliyetler yeryüzü üzerindeki sınırlı doğal kaynakların tahrip etmekte ve birtakım çevresel probleme neden olmaktadır. Dünya nüfusundaki artış beraberinde hızlı ve çarpık kentleşmelere neden olmakta, dolayısıyla sınırlı doğal kaynaklar üzerinde olumsuz bir etki yaratmaktadır. Bu açıdan söz konusu değişimlerin belirlenmesi doğal kaynakların korunması, yönetimi ve planlanması ile ilgili küresel ve yerel ölçekli birçok çalışma için büyük önem arz etmektedir.

Değişimin belirlenmesi yeryüzünün belirli aralıklarla gözlemlenmesini ve elde edilen verilerin analizini gerektirmektedir. Uzaktan algılama teknolojileri belirli bir bölgeyi farklı zaman aralıklarında, farklı konumsal ve spektral çözünürlüklerde görüntülenebilmektedir. Sahip olduğu bu özellikler uzaktan algılanmış görüntüleri yeryüzü ile ilgili yapılan çalışmalarda kullanılan önemli bir veri kaynağı durumuna getirmiştir. Uzaktan algılanmış görüntülerin sınıflandırılması ile yeryüzünün genel arazi kullanımı veya arazi örtüsü belirlenebilmekte ve elde edilen sonuçlar analiz edilebilmektedir.

Uzaktan algılanmış görüntülerin sınıflandırılması işlemi kullanılan bir takım sınıflandırma algoritmaları ile görüntü üzerindeki piksellerin yansıtma ve parlaklık değerlerine göre

belirlenen arazi örtüsü veya arazi kullanımı gruplarından birine atanması olarak ifade edilebilir (Lillesand et al., 2008). Günümüze kadar uydu görüntülerinin sınıflandırılması ve tematik haritaların üretilmesi amacıyla birçok algoritma geliştirilmiş ve uygulamalarda kullanılmıştır (Jain et al., 2000; Lu and Weng, 2007). Bu algoritmalar arasında en yakın komşuluk ve en çok benzerlik algoritmaları uygulamalarda en çok kullanılan metotlardır. Sınıflandırmada kullanılan algoritmalar sınıflandırma sonucunu dolayısıyla üretilen tematik haritaların doğruluğunu direkt olarak etkilemektedir (Foody, 2004). Bu nedenle uzaktan algılama literatüründe tematik haritaların doğruluğunu arttırmaya yönelik yeni metotların geliştirilmesi önemli bir çalışma konusu olmuştur (De Leeuw et al., 2006; Kavzoglu, 2009). Cleary and Trig, (1995) tarafından önerilmiş örnek tabanlı sınıflandırıcılardan entropik uzaklık ölçütünü kullanan K-star algoritması sınırlı sayıda test edilmiş bir yöntemdir. Yöntem kullanıcı tarafından belirlenmesi gereken tek bir parametreye sahip olup matematiksel olarak kolay ifade edilmektedir.

Bu çalışmada, K-star algoritması ile 2009 ve 2010 tarihinde kaydedilen Landsat TM uydu görüntüleri kullanılarak Kocaeli ilçesinin genel arazi kullanımı ve arazi örtüsünün tespit edilerek tematik haritaların üretilmesi amaçlanmıştır. Yöntemin genel sınıflandırma doğruluğu, en çok benzerlik ve en yakın komşuluk yöntemleri ile elde edilen sınıflandırma doğrulukları ile karşılaştırılmıştır. Elde edilen sınıflandırma doğrulukları arasındaki farkların istatistiksel anlamlılığı Z testi kullanılarak analiz edilmiştir.

2. ÇALIŞMA ALANI VE KULLANILAN VERİ

Yaklaşık yüzölçümü 3.623 km² olan Kocaeli ili ülkemizin en önemli endüstri ve sanayi bölgelerinden birisidir. Doğusunda Sakarya, batısında İstanbul, güneyinde Yalova illeri ile sınırlanan sanayi kenti Asya ile Avrupa kıtaları arasında önemli kara ve demiryolu güzergâhlarının kesiştiği bir yerde bulunmaktadır (Şekil 1). Kocaeli ilinin 1980 yılındaki genel nüfus sayımında 596.899 olan nüfusu, 2010 yılında yapılan adrese dayalı nüfus kayıt sistemine göre 1.560.138'e yükselmiştir (www.tuik.gov.tr).



Şekil 1. Çalışma alanı (Kocaeli) il sınırları.

Bu çalışmada Kocaeli ilinin genel arazi kullanımı ve arazi örtüsünün belirlenmesi amacıyla 30.09.2009 ve 31.07.2010 tarihlerinde kaydedilen Landsat TM uydu görüntüleri kullanılmıştır. Uydu görüntülerinin geometrik olarak rektifikasyonunda bölgeye ait 1/25.000 ölçekli haritalardan yararlanılmış ve görüntüler birinci derece Afın dönüşümü kullanılarak UTM projeksiyon sistemine dönüşümü gerçekleştirilmiştir. Uydu görüntülerinin yeniden örneklenmesinde orijinal değerler korunduğundan en yakın komşuluk yöntemi tercih edilmiştir.

3. SINIFLANDIRMA YÖNTEMLERİ

3.1. Örnek Tabanlı Öğrenme ve K-star Algoritması

Örnek tabanlı öğrenme algoritmaları literatürde en yaygın kullanımı olan öğrenme algoritmalarıdır (Okamoto and Yugami, 2003). Örnek tabanlı metotlar test veri setindeki öznitelik bilgisi bilinmeyen bir pikselin sınıflandırılması işlemini veri tabanında saklanan ve önceden sınıflandırılmış eğitim veri setindeki pikseller ile karşılaştırarak gerçekleştirirler (Aha et al., 1991; Aha, 1992; Wilson and Martinez, 2000). Bu karşılaştırma işleminde temel kabul benzer piksellerin benzer sınıflandırmalara sahip olacağıdır (Piramuthu and Sikora, 2009). Burada esas olan benzer pikseller ve benzer sınıflandırmanın nasıl belirleneceğidir. Örnek tabanlı öğrenciler iki piksel arasındaki benzerliğin belirlenmesinde bazı uzaklık veya benzerlik fonksiyonlarından yararlanırlar. Bu uzaklık fonksiyonu hangi pikselin girdi vektörüne en yakın olduğuna karar verilmesinde

kullanılmaktadır (Wilson and Martinez, 2000). Örnek tabanlı öğrenmede yeni bir piksel ile çevresindeki pikseller arasındaki uzaklık ölçülür. Öklit, Mahalanobis, Quadratic ve Ki-Kare gibi uzaklık fonksiyonları pikseller arasındaki benzerliğin veya spektral uzaklığın belirlenmesinde en sık kullanılan fonksiyonlardır (Diday, 1974; Michhalski et al., 1981; Nadler and Smith, 1993). En yakın komşuluk ve k-en yakın komşuluk algoritmaları literatürde yaygın olarak kullanılan en basit örnek tabanlı öğrenme algoritmalarındandır (Cover and Hart, 1967).

Uzaktan algılanmış görüntülerin sınıflandırılmasında son yıllarda kullanılan K-star sınıflandırıcısı iki piksel arasındaki spektral uzaklığın veya iki piksel arasındaki benzerliğin belirlenmesinde bilgi teorisine dayanan entropik uzaklık ölçüsünü kullanmaktadır (Cleary and Trig, 1995). Bilgi teorisinde, entropi rastgele değişkenler ile ilgili belirsizliğin ölçüsü olarak tanımlanmıştır (Shannon, 1948). Pikseller arasındaki uzaklığa ilişkin bu öngörü bir pikselin başka bir piksele dönüşümündeki karmaşıklık olarak tanımlanır. Bu karmaşıklığın hesaplanması iki adımda gerçekleştirilir. İlk olarak, özellikten özelliğe dönüşüm olarak tanımlanan sonlu sayıda dönüşümler kümesi belirlenir. Daha sonra bir özelliğin (a) diğer bir özelliğe (b) dönüşümü, (a)'dan başlayıp (b)'de sonlanan sınırlı dönüşümler dizisi olarak tanımlanır (Cleary and Trig, 1995). K-star algoritması iki özelliği birbirine bağlayan en kısa uzaklık olarak Kolmogorov mesafesini dikkate almaktadır. Bu durumda K-star uzaklığı, iki özellik arasındaki tüm olası dönüşümlerin toplamı olmaktadır (Piramuthu and Sikora, 2009). Olasılık fonksiyonu P^* , t dönüşümleri ile a özelliğinden b özelliğine olan tüm yolların olasılığı olarak,

$$P^*(b/a) = \sum_{t \in P: t(a)=b} \bar{p}(t)$$

(1)

belirlenir. Bu durumda K-star (K^*) fonksiyonu,

$$K^*(b/a) = -\log_2 P^*(b/a)$$

(2)

şeklinde ifade edilir.

K-star fonksiyonu sayısal veya sembolik öznitelik değerine sahip veri setlerinin sınıflandırılmasında kullanılabilir. Fonksiyonun sayısal sınıf değerleri için kullanımında x_0 , sembolik sınıf değerleri için kullanımında s parametreleri eğitim öncesinde kullanıcı tarafından belirlenmektedir. Her iki durumda da olasılık dağılımına dahil olan örnek sayısı, en yakın komşuluk dağılımı olan 1 değeri ile tüm örneklerin eşit ağırlığa sahip olduğu N değeri aralığında değişir (Cleary and Trig, 1995). Herhangi bir P^* fonksiyonu için etkili örnek sayısı,

$$n_0 \leq \frac{\left(\sum_b P^*(b/a) \right)^2}{\sum_b P^*(b/a)^2} \leq N$$

(3)

şeklinde ifade edilebilir. Burada N eğitim örneklerinin tümünün sayısını, n_0 ise eğitim örneklerinden a örneğine en kısa mesafedeki örneklerin sayısını gösterir. K-star algoritması x_0 ve s parametreleri için a değerini n_0 ve N arasındaki bir değer seçilmesiyle belirler. Böylece seçilen n_0 en yakın komşuluk algoritmasını ve seçilen N eşit ağırlıklı örneklerin sayısını verir. Belirlenen bu sayıların uygunluğu harmanlama parametresi (b) kullanılarak belirlenir. Harmanlama parametresi $b = \%0$ (n_0 için) ve $b = \%100$ (N için) arasında değişmektedir.

3.2. En Çok Benzerlik Sınıflandırıcısı

En çok benzerlik (EÇB) yöntemi uzaktan algılanmış görüntülerin sınıflandırılmasında yaygın olarak kullanılan istatistiksel bir sınıflandırma yöntemidir. Yöntem sınıflandırma işleminin gerçekleştirilmesinde ortalama değer, varyans ve kovaryans gibi istatistiksel değerleri dikkate almaktadır. EÇB sınıflandırıcısı, eğitim örneklerini oluşturan sınıflar için olasılık fonksiyonları hesaplar ve her bir pikselin hangi sınıfa daha yakın olduğunu tespit eder. Bir piksele ait sınıf etiketi hesaplanan olasılıklara göre, pikselin en yüksek olasılıklı sınıfa atanması ile belirlenir. EÇB yönteminde, eğitim verilerinin ve sınıfları oluşturan piksellerin normal dağılımda olduğu kabul edilir (Campbell, 1996). Yöntem sınıflandırmada, her sınıf için oluşturacak varyans-kovaryans matris değerlerini de dikkate alarak, örnek piksellerin özelliklerindeki dağılımlarını dikkate almaktadır (Mather, 1999). Pikselin p sayıdaki bant değerlerini içeren x vektörünün bir k sınıfında olma olasılığı aşağıdaki şekilde hesaplanır.

$$P(x) = 2\pi^{-0.5p} |S_i|^{-0.5} \exp\left[-0.5(y^1 S_i^{-1} y)\right] \quad (4)$$

Bu eşitlikte; $P(x)$, olasılık değerini; S_i i . sınıf için varyans-kovaryans matrisini; $| |$ matrisin determinantını ifade etmektedir.

3.3. En Yakın Komşuluk Sınıflandırıcısı

Bu yöntem, sınıfları belli olan bir örnek kümesindeki gözlem değerlerinden yararlanarak, örneğe katılacak yeni bir gözlemin hangi sınıfa ait olduğunu belirlemek amacıyla kullanılır. Bu algoritmalar eğitim verisinden en çok bezeyen tek bir örneği bulmak için bir takım alan özellikli uzaklık fonksiyonları kullanılır. Bulunan örnek yeni bir örneğin sınıflandırılması için kullanılır. En yakın komşuluk algoritması eğitim verisindeki piksellerin her birinin, öznelik değeri bilinmeyen test verisindeki bir piksele olan uzaklıkların hesaplanması ve en yakın uzaklığa sahip k sayıda gözlemin seçilmesi esasına dayanmaktadır. Yöntem uzaklıkların hesaplanmasında, i ve j pikselleri için Eşitlik (5)'te formülasyonu verilen Öklit uzaklığını kullanmaktadır.

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (5)$$

4. UYGULAMA

Her iki görüntü üzerinden yapılan ön çalışmalar neticesinde Kocaeli ilinin mevcut arazi örtüsü ve kullanımını temsil eden

altı temel sınıfın (su, iğne yapraklı orman, geniş yapraklı orman, bozkır, toprak-taş ve şehir) mevcut olduğu tespit edilmiştir. Sınıflandırmada kullanılmak üzere örnekleme alanları mevcut haritalar, hava fotoğrafları ve araziden el-GPS'i ile toplanan veriler yardımıyla oluşturulmuştur. Belirlenen örnekleme alanları uydu görüntüleri ile birleştirilerek sınıflandırma işleminde kullanılmak üzere her iki uydu görüntüsü için rastgele örnekleme prensibi esas alınarak eğitim ve test veri setleri oluşturulmuştur. Hesaplanan genel sınıflandırma doğruluklarının karşılaştırılmasında objektifliğin sağlanması amacıyla eğitim ve test veri setleri belirlenirken eşit sayıda örnek seçilmiştir. Sonuç olarak belirlenen altı arazi örtüsü sınıfı için eğitim verisi olarak 3.600 piksel (her bir sınıf için 600 piksel), test verisi olarak toplam 3.000 piksel (her bir sınıf için 500 piksel), her iki görüntü için ayrı ayrı seçilmiştir.

Çalışmada K-star algoritmasının sınıflandırma performansının en çok benzerlik ve en yakın komşuluk algoritmalarının performansları ile karşılaştırılmasında öncelikle sınıflandırma sonucunda elde edilen genel doğruluklar kullanılmıştır. Bu değerlendirmenin yanı sıra sınıflandırma performansları arasındaki farkların istatistiksel olarak anlamlılığı Z testi kullanılarak analiz edilmiştir. Z testi, sınıflandırma sonucu elde edilen Kappa değerlerinin karşılaştırılması amacıyla kullanılan istatistiksel bir testtir. İki farklı sınıflandırma yöntemi veya iki farklı eğitim seti için elde edilen genel sınıflandırma doğruluğu arasındaki farkın istatistiksel olarak anlamlı olup olmadığını tespitinde hesaplanan Kappa değerlerinin karşılaştırılması esasına dayanmaktadır (Foody, 2004). Z istatistik değeri aşağıdaki şekilde hesaplanarak Kappa değerleri arasındaki farkın anlamlılığı test edilir.

$$Z = \frac{K_1 - K_2}{\sqrt{\sigma(K_1) + \sigma(K_2)}} \quad (6)$$

Bu eşitlikte K_1 ve K_2 karşılaştırılan iki Kappa değerini, $\sigma(K_1)$ ve $\sigma(K_2)$ ise Kappa değerleri için hesaplanan varyans değerlerini ifade etmektedir. 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 sonucu 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 göstermektedir.

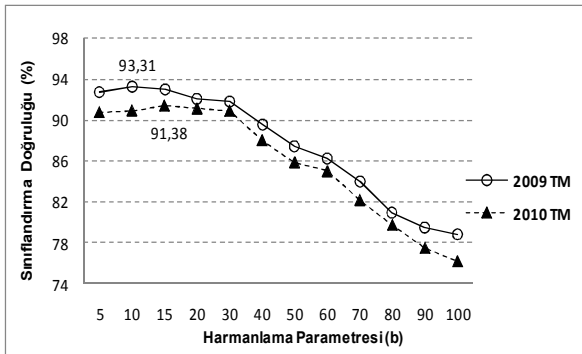
2009 ve 2010 tarihli Landsat TM görüntüleri için rastgele örnekleme prensibinden hareketle oluşturulan eğitim veri setleri kullanılarak en çok benzerlik (EÇB) yöntemi için sınıflandırmaya esas olacak modeller oluşturulmuş ve elde edilen modellerin geçerlilikleri test veri setleri kullanılarak analiz edilmiştir. En çok benzerlik yöntemi için elde edilen genel sınıflandırma doğrulukları ve Kappa değerleri Tablo 1'de gösterilmiştir. Tablodan da görüleceği üzere 2009 tarihli TM görüntüsü için genel doğruluk $\%92,75$ ve Kappa değeri 0,91 olarak hesaplanmıştır. Diğer taraftan 2010 tarihli uydu görüntüsü için genel doğruluk $\%90,75$ ve Kappa değeri 0,89 olarak belirlenmiştir.

Tablo 1. Kullanılan yöntemleri için elde edilen sınıflandırma sonuçları.

| Yöntemler | 2009 Landsat TM | | 2010 Landsat TM | |
|-----------|--------------------|-------|--------------------|-------|
| | Genel Doğruluk (%) | Kappa | Genel Doğruluk (%) | Kappa |
| EÇB | 92,75 | 0,91 | 90,75 | 0,89 |
| EYK | 91,72 | 0,90 | 90,41 | 0,88 |
| K-star | 93,31 | 0,92 | 91,38 | 0,90 |

En yakın komşuluk (EYK) algoritması ile sınıflandırma işlemi Matlab yazılımında hazırlanan bir program ile gerçekleştirilmiş ve test verileri için doğruluk analizi gerçekleştirilmiştir. Sınıflandırma sonucunda 2009 tarihli uydu görüntüsü için hesaplanan genel doğruluk %91,72, 2010 görüntüsü için %90,4'dir. Kappa değerleri ise sırasıyla 0,90 ve 0,88 olarak hesaplanmıştır.

Örnek tabanlı öğrenme algoritmalarından K-star algoritması pikseller arasındaki benzerliğin belirlenmesi için entropik uzaklık fonksiyonu kullanılır. Bu algoritma ile sınıflandırmaya esas model oluşumunda harmanlama parametresi (b) kullanıcı tarafından belirlenmektedir. Çalışmada kullanılan Landsat TM görüntülerinin K-star algoritması ile sınıflandırılması için model oluşumunda en uygun harmanlama parametresinin belirlenmesinde eğitim ve test veri setleri kullanılmıştır. Şekil 2'de her iki uydu görüntüsü için seçilecek harmanlama parametresindeki değişimler ve bu parametre değerlerine karşılık gelen genel sınıflandırma doğrulukları gösterilmektedir. Şekilden de görüleceği üzere her iki uydu görüntüsü için harmanlama parametresi 1-30 aralığındaki değerler ile %90'ın üzerinde genel sınıflandırma doğruluğu elde edilmektedir. Bu aralıkta 2009 TM görüntüsü için parametre değeri 10 olduğunda en yüksek sınıflandırma doğruluğuna (%93,31) ulaşıırken, 2010 TM görüntüsü için en yüksek sınıflandırma doğruluğunun (%91,38) elde edildiği parametre değeri 15 kullanılmıştır.

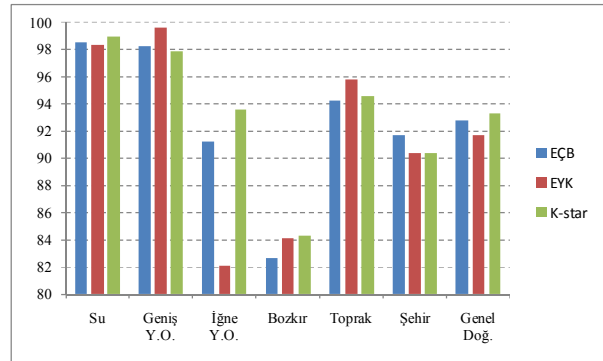


Şekil 2. Landsat TM görüntüleri için harmanlama parametresi ve genel sınıflandırma doğruluğundaki değişimler.

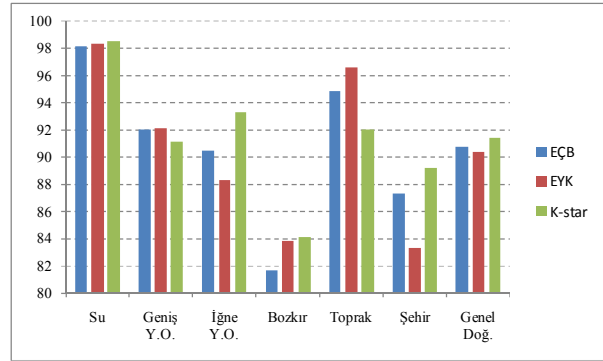
Çalışmada kullanılan üç sınıflandırma metodu için elde edilen sınıflandırma doğrulukları karşılaştırıldığında her iki uydu görüntüsü için en yüksek sınıflandırma doğruluğunun K-star sınıflandırıcısı ile elde edildiği görülmektedir. Sınıflandırma sonucu hesaplanan sınıf doğrulukları analiz edildiğinde Şekil 4'ten de görüleceği üzere toprak sınıfı için en yakın komşuluk yöntemi ile daha yüksek doğruluğa ulaşıldığı görülmektedir. Diğer taraftan özellikle iğne yapraklı orman ve bozkır alanlarının sınıflandırılmasında en düşük doğrulukların en

yakın komşuluk yöntemi ile elde edildiği ifade edilebilir. Şekilden de görüleceği üzere spektral olarak benzer piksellere sahip iğne ve geniş yapraklı orman sınıflarının sınıflandırılmasında K-star sınıflandırıcısı ile daha tutarlı sonuçlar elde edilmiştir.

Uygulamada K-star sınıflandırıcısının sınıflandırma performansı en çok benzerlik ve en yakın komşuluk yöntemleri ile genel sınıflandırma doğrulukları esas alınarak karşılaştırılmasının yanı sıra sınıflandırma performanslarındaki farklılıklarının istatistiksel olarak anlamlılığı Z testi ile analiz edilmiştir. Bu amaçla her iki uydu görüntüsü için oluşturulan hata matrisleri kullanılarak hesaplanan Z istatistik değerleri Tablo 2'de gösterilmiştir. Tablodan da görüleceği üzere K-star ile en çok benzerlik (EÇB) yöntemi ve K-star ile en yakın komşuluk (EYK) yöntemi için yapılan istatistik testi sonuçları $Z=1,96$ kritik değerinden büyüktür. Test sonuçları K-star yöntemi ile diğer iki yöntemler arasındaki sınıflandırma performansları arasındaki farklarının istatistiksel olarak anlamlı olduğunu göstermektedir. Elde edilen bu sonuç çalışmada kullanılan veri setleri için K-star algoritmasının EÇB ve EYK yöntemlerine göre daha yüksek doğruluklu sınıflandırma sonuçları ürettiğini destekler niteliktedir.



(a)



(b)

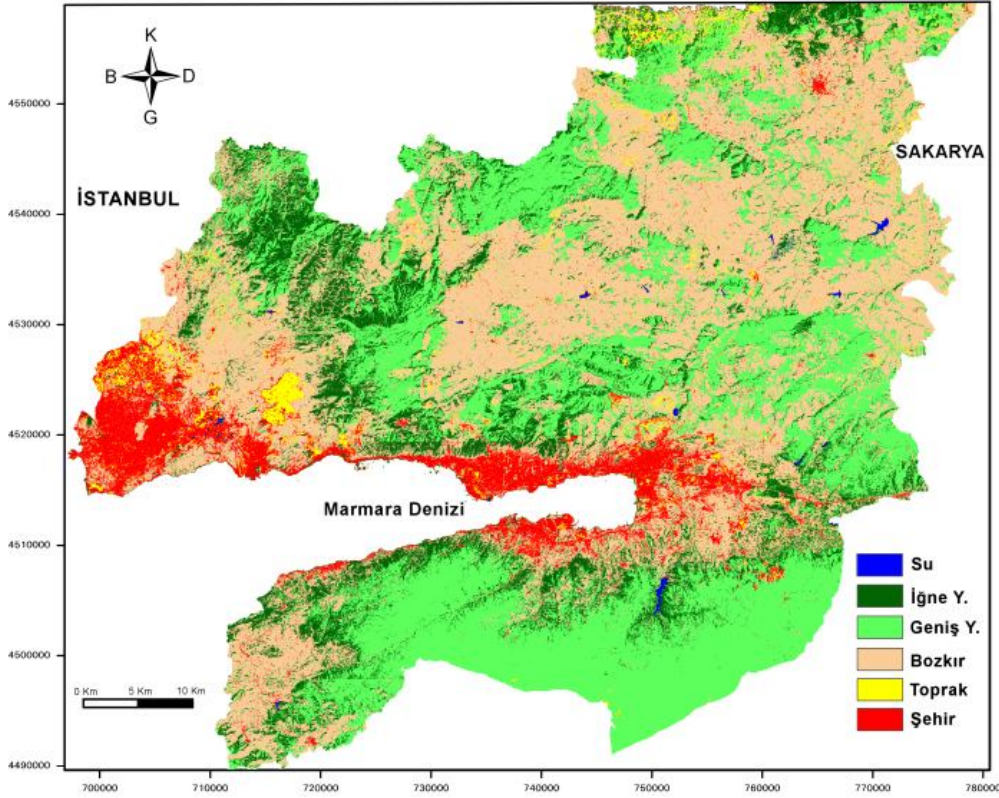
Şekil 3. En çok benzerlik, en yakın komşuluk ve K-star sınıflandırıcıları kullanılarak (a) 2009 Landsat TM, (b) 2010 Landsat TM görüntüleri için elde edilen sınıflandırma doğrulukları.

Tablo 2. K-star, EÇB ve EYK yöntemlerinin performans analizinde hesaplanan Z test değerleri ($Z=1,96$ kritik değer).

| | 2009 Landsat TM | 2010 Landsat TM |
|--------------|-----------------|-----------------|
| K-star ~ EÇB | 2,05 | 2,09 |

Çalışmada kullanılan tüm sınıflandırıcılar için eğitim ve test verileri kullanılarak sınıflandırmaya esas olacak modeller ayrı ayrı oluşturularak uydu görüntülerine uygulanmış ve çalışma alanının genel arazi örtüsünü gösteren tematik haritalar üretilmiştir. Söz konusu tematik haritalara örnek olarak 2010 tarihli güncel Landsat TM görüntüsünün K-star sınıflandırıcısı ile elde edilen Şekil 3'de gösterilmiştir. Tematik harita incelendiğinde Kocaeli ili genelinde şehirleşmenin özellikle

Marmara denizini çevreleyen kıyı bölgelerinde yoğunlaştığı görülmektedir. Bu durum sınırlı kaynaklar arasında olan kıyı bölgelerinin yoğun bir şehirleşme baskısı altında olduğunu göstermektedir. Çalışma alanındaki iğne yapraklı orman alanlarının özellikle şehrin Karadeniz'e kıyısı buluna Kuzey-Doğu kesiminde ve İstanbul il sınırına yakın Kuzey-Batı kesiminde yoğunlaştığı görülmektedir. Kocaeli ilinin güneyinde ise geniş yapraklı orman alanlarının sahip olduğu görülmektedir.



Şekil 3. K-star algoritması ile 2010 tarihli Landsat TM görüntüsünün sınıflandırılması sonucu elde edilen tematik harita.

5. SONUÇLAR

Uzaktan algılamada sınıflandırma probleminin çözümü ve yeryüzünü temsil eden tematik haritaların üretilmesi amacıyla farklı matematiksel algoritmaları kullanan birçok sınıflandırma yöntemi geliştirilmiştir. Literatürde sınıflandırma yöntemlerinin farklı veri setleri ve sınıflandırma probleminin çözümündeki performansları ve birbirlerine göre farklılıklarının analizi ile ilgili yoğun çalışmalar yürütüldüğü görülmektedir.

Bu çalışmada sınıflandırma yöntemlerinden entropik uzaklık fonksiyonunu kullanan K-star algoritmasının uzaktan algılanmış görüntülerin sınıflandırılmasındaki kullanımı iki farklı veri seti kullanılarak incelenmiştir. Yöntem sınıflandırma öncesinde harmanlama parametresi olarak bilinen tek bir parametrenin belirlenmesini gerektirmektedir. Çalışmada kullanılan iki farklı veri seti için söz konusu parametre değişimlerine karşılık elde edilen sınıflandırma doğrulukları incelenmiştir. Parametre değişimleri analiz

edildiğinde her iki veri seti içinde harmanlama parametresinin 0-30 aralığında alacağı değerler için %90'ın üzerinde sınıflandırma doğruluğu elde edilmiştir. En yüksek sınıflandırma doğruluğunun elde edildiği parametre değerleri 2009 tarihli görüntü için 10; 2010 tarihli görüntü için 15 olarak belirlenmiş ve bu parametrelerle sınıflandırmaya esas modeller oluşturulmuştur.

K-star algoritmasının sınıflandırma performansı literatürde en sık kullanılan iki yöntemin (en çok benzerlik ve en yakın komşuluk) sınıflandırma performansı ile karşılaştırılmıştır. 2009 tarihli Landsat TM görüntüsünün K-star ile sınıflandırılması sonucunda genel sınıflandırma doğruluğu yaklaşık %94 olarak hesaplanırken, en çok benzerlik (EÇB) yöntemi ile %93, en yakın komşuluk (EYK) ile %91 olarak hesaplanmıştır. 2010 Landsat TM görüntüsü için K-star, EÇB ve EYK ile hesaplanan genel sınıflandırma doğrulukları sırasıyla yaklaşık olarak %92, %91 ve %90 olarak hesaplanmıştır. Yöntemler için söz konusu performans değişikliklerinin istatistiksel anlamlılığı Z testi kullanılarak analiz edilmiştir. İstatistiksel analizde test veri seti için

hesaplanan hata matrisleri kullanılmış ve her iki uydu görüntüsü için K-star yönteminin diğer iki yöntemle olan performans farklılığı analiz edilmiştir. İstatistiksel test sonuçları her iki uydu görüntüsünün sınıflandırma probleminde K-star ile elde edilen genel sınıflandırma doğruluğunun, EÇB ve EYK yöntemleri kullanılarak elde edilen sınıflandırma doğruluklarından farklı olduğunu göstermiştir. Bu sonuç K-star algoritması ile sınıflandırma doğruluğunda elde edilen yaklaşık %2'lik iyileşmenin istatistiksel olarak anlamlı olduğunu ve bu çalışmada kullanılan veri setleri için EÇB ve EYK yöntemlerine göre daha doğru sonuçlar ürettiğini desteklemektedir. Yapılan performans analizleri ve elde edilen istatistiksel test sonuçları kullanıcı tarafından belirlenen tek bir parametreye sahip K-star algoritmasının uzaktan algılanmış görüntülerin sınıflandırmasında etkili bir yöntem olabileceğini göstermektedir.

KAYNAKLAR

- Aha, D.W., 1992. Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies*, 36(2), pp. 267-287.
- Aha, D.W., Kibler, D., Albert, M.K., 1991. Instance-based learning algorithms. *Machine Learning*, 6(1), pp. 37-66.
- Campbell, J.B., 1996. *Introduction to Remote Sensing*, Second Edition, The Guilford Press, New York.
- Cleary, J.G., Trigg, L.E., 1995. K*: An instance-based learner using an entropic distance measure. *Proceedings of the 12th International Conference on Machine Learning*, Tahoe City, California, USA, pp.108-114.
- Cover, T.M., Hart, P.E., 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), pp. 21-27.
- De Leeuw, J., H. Jia, L. Yang, X. Liu, K. Schmidt, A.K. Skidmore, 2006. Comparing accuracy assessments to infer superiority of image classification methods. *International Journal of Remote Sensing*, 27(1), pp. 223-232.
- Diday, E., 1974. Recent progress in distance and similarity measures in pattern recognition. *Proceedings of the Second International Joint Conference on Pattern Recognition*, Copenhagen, Denmark, pp. 534-539.
- Foody, G.M., 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 70(5), pp. 627-633.
- Jain, A.K., Duin, R.P.W., Mao, J.C., 2000. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), pp. 4-37.
- Kavzoglu, T., 2009. Increasing the accuracy of neural network classification using refined training data. *Environmental Modelling & Software*, 24(7), pp. 850-858.
- Lillesand, T.M., Kiefer, R.W., Chipman, J.W., 2008. *Remote Sensing and Image Interpretation*. Sixth Edition, John Wiley & Sons, New York.
- Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5), pp. 823-870.
- Mather, P.M., 1999, *Computer Processing of Remotely Sensed Images*, Second Edition, John Wiley & Sons, Chichester.
- Michalski, R.S., Stepp, R.E., Diday, E., 1981. *A Recent Advance in Data Analysis: Clustering Objects into Classes Characterized by Conjunctive Concepts*. In Progress in Pattern Recognition, Kanal, L.N. and A. Rosenfeld (Eds.), North-Holland, New York, pp. 33-56.
- Nadler, M., Smith, E. P., 1993. *Pattern Recognition Engineering*. John Wiley & Sons, New York, 608 p.
- Okamoto, S., Yugami, N., 2003. Effects of domain characteristics on instance-based learning algorithms. *Theoretical Computer Science*, 298(19), pp. 207-233.
- Piramuthu, S., Sikora, R.T., 2009. Iterative feature construction for improving inductive learning algorithms. *Expert Systems with Applications*, 36(2), pp. 3401-3406.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27, pp. 379-423.
- Wilson, D.R. and Martinez, T.R., 2000. Reduction techniques for instance-based learning algorithms. *Machine Learning*, 38(3), pp. 257-286.