

Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması**

Özlem Akar*, Oğuz Güngör

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

Özet

Web Yayın: 12 Mart 2013

Cilt: 1
Sayı: 2
Sayfa: 139 - 146
Kasım 2012

Rastgele Orman (RO) algoritması en başarılı sınıflandırma yöntemlerinden biri olarak bilinir. Doğası gereği çok farklı disiplinlere hitap etmesinden dolayı, RO farklı alanlarda çalışan araştırmacıların dikkatini çekmektedir. Bu çalışma, farklı konumsal çözünürlüğe ve karakteristiğe sahip çok bantlı uydu görüntüleri kullanarak RO algoritmasının performansını incelemeyi amaçlamaktadır. Kullanılan uydu görüntüleri dört bantlı Ikonos ve QuickBird görüntüleridir. 2005 ve 2008 yıllarında elde edilen QuickBird görüntüleri sırasıyla hem kentsel hem de kırsal alanları kapsarken, 2003 yılında alınan Ikonos görüntüsü, özellikle kentsel alanı içermektedir. Ayrıca, 2005 yılında alınan QuickBird görüntüsü rüzgarlı havanın yol açtığı dalgalar nedeniyle Karadeniz üzerinde gürültülü görüntüler içermektedir. RO'nun performansını değerlendirmek için sınıflandırma sonuçları, Gentle AdaBoost (GAB), En Çok Benzerlik (EÇB) ve Destek Vektör Makineleri (DVM) algoritmalarından elde edilen sonuçlarla karşılaştırılmıştır. Elde edilen sonuçlar RO'nun diğer yöntemlerden daha yüksek sınıflandırma doğruluğu verdiğini göstermektedir. Kentsel alan üzerinde çekilen Ikonos görüntüsüne ait sonuçlar, RO algoritmasının, DVM'den %10 daha yüksek sınıflandırma doğruluğu verdiğini, GAB algoritmasının ise en düşük sınıflandırma doğruluğuna sahip olduğunu (RO'dan %14 daha düşük) göstermektedir. Kırsal alan üzerinde alınan QuickBird görüntüsüne (2008 yılında alınan) ait sonuçlar diğer yöntemlerden elde edilen sonuçlarla karşılaştırıldığında RO'nun daha iyi sonuç verdiği görülmüştür. Gürültüye benzer görüntüler içeren QuickBird görüntüsü için de RO'nun, DVM'den yaklaşık %11 daha yüksek sınıflandırma doğruluğu verdiği gözlenmiştir.

Anahtar Sözcükler

Görüntü sınıflandırma, Rastgele orman algoritması, Doğruluk analizi, Arazi kullanımı.

Abstract

Classification of multispectral images using Random Forest algorithm

Pub. Online: 12 March 2013

Volume: 1
Issue: 2
Page: 139 - 146
November 2012

Random Forest (RF) algorithm is known to be one of the most efficient classification methods. Due to its inherent interdisciplinary nature, it draws researchers from different backgrounds. This study aims at investigating the performance of RF algorithm using multispectral satellite images having different spatial resolutions and scene characteristics. The satellite images used include Ikonos and QuickBird images with four multispectral bands. Ikonos image taken in 2003 covers mainly urban area, whereas QuickBird images acquired in 2005 and 2008 covers both urban and rural areas, respectively. QuickBird image taken in 2005 also contains noisy patterns over Black Sea due to waves resulting from windy weather. To evaluate the performance of RF, the classification results are compared with the results obtained from Gentle AdaBoost (GAB), Support Vector Machine (SVM) and Maximum Likelihood Classification (MLC) algorithms. Preliminary results indicate that RF gives higher classification accuracies than other methods. For Ikonos image over urban area, the results show that RF algorithm gives 10% higher classification accuracy than SVM, whereas GAB algorithm has the lowest classification accuracy (14 % lower than RF). For QuickBird image (taken in 2008) of rural area, RF gives the best result compared to the others. Also, for QuickBird image containing noisy pattern, RF has around 11% higher overall accuracy than SVM.

Keywords

Image classification, Random forest algorithm, Accuracy assessment, Land use.

** Hakemli değerlendirme sürecinden geçerek, Jeodezi ve Jeoinformasyon Dergisi'nde yayına kabul edilen ve asıl yazı dili İngilizce olan "Classification of multispectral images using random forest algorithm" başlıklı eserin Türkçe diline çevirisidir. Çeviren; Özlem Akar.

*Sorumlu Yazar: Tel: +90 (462) 3772764 Fax: +90 (462) 328 0918

E-posta: oerden@ktu.edu.tr (Akar Ö.), oguzgungor@gmail.com (Güngör O.)

1. Giriş

Görüntü sınıflandırma, görüntüde her bir piksel konumundaki gri değerlerini kullanarak bu bilgileri anlamlı arazi örtüsü bilgisine dönüştürme işlemidir. Başka bir deyişle görüntü sınıflandırma, görüntü üzerindeki pikselleri spektral alandaki istatistiksel karar kuralları ya da mekânsal alandaki mantıksal karar kuralları kullanarak çok sayıda sınıfa atama işlemidir. Spektral alan, görüntüdeki piksellerin spektral değerlerini esas alan karar kurallarını kullanmaktadır. Mekânsal alandaki karar kuralları ise piksellerin komşuluklarını, oluşturdukları şekil, doku ve desen gibi konumsal içeriği esas almaktadır (Gao 2009).

Literatürde sınıflandırma işlemi için en yaygın kullanılan yaklaşımlar piksel tabanlı ve obje tabanlı yaklaşımlardır. Piksel tabanlı sınıflandırma yaklaşımları sadece her piksel konumunda spektral vektörler olarak da adlandırılan spektral bilgiyi kullanır ve konumsal içerik göz ardı edilir. En yaygın kullanılan piksel tabanlı yaklaşımlardan biri de En Çok Benzerlik (EÇB) sınıflandırma algoritmasıdır. EÇB yöntemi her bir bantta her bir sınıf için görüntü verisini normal dağılımda kabul eder. EÇB işleminde verilen bir piksel, belirli bir sınıfa ait olma olasılığına sahiptir. Bundan dolayı, her bir pikselin olasılığı hesaplanır ve her piksel en yüksek olasılığa sahip olan sınıfa atanır (ENVI 2005).

Buna karşın obje tabanlı yaklaşımlar daha güvenilir sonuçlar elde etmek için pikselleri, onların konumsal özelliklerini de esas alan piksel grupları olarak ele alır. Objeye tabanlı sınıflandırma yöntemleri, spektral değer yanında şekil ve doku gibi farklı obje özelliklerini de kullanır. Bu yöntemler şekil, yoğunluk ve ölçek faktörü gibi kriterler ile elde edilen piksel gruplarını temsil eden objelere göre görüntüyü parçalara ayırır. Sonrasında bu parçalar sınıflandırılır (Dronova vd. 2011). Bu yaklaşımın temelinde bulanık mantık teorisi yer alır. Başka bir deyişle obje, farklı üyelik değerleri ile birden fazla sınıfa atanabilir (Matinfar 2007).

Klasik piksel ve obje tabanlı yaklaşımlara alternatif olarak uydu görüntülerinden daha doğru ve güvenilir bilgi elde etmek için çeşitli öğrenme tabanlı algoritmalar geliştirilmektedir. En yaygın kullanılan öğrenme tabanlı algoritmalar Rastgele Orman (RO), Torbalama, Hızlandırma, Karar Ağacı, Yapay Sinir Ağları, Destek Vektör Makinesi (DVM) ve K-En Yakın Komşuluk'tur. Bu algoritmalar makine öğrenme metodları olarak da bilinir. İstatistiksel yöntemlerin aksine makine öğrenme yöntemleri verinin dağılımı ile ilgili herhangi bir varsayıma dayanmadığı için parametrik değildir. Bunlar veri temelli yöntemlerdir ve tahmin edici ve buna karşılık gelen cevaplar arasındaki ilişkileri öğrenirler (Breiman 2001). Makine öğrenme yöntemleri, yeteri büyüklükteki veri setlerinden ve parametrelerden faydalanarak, veri için en uygun modeli girdi veriden elde edilen karar kurallarını kullanarak bulmayı amaçlarlar.

Son yıllarda yaygın olarak kullanılan makine öğrenme sınıflandırıcılarından olan DVM, sınıfları birbirinden ayırmak için maksimum marjli doğrusal bir ayırt edici fonksiyon bulmayı amaçlar. Doğrusal olarak ayrılabildikleri daha yüksek boyutlu başka bir uzaya taşınır ve sınıflandırma işlemi o uzayda gerçekleşir (Kaban ve Diri 2008). Aslında DVM, çoklu sınıf problemleri kompleks optimizasyon gerektirdiği için ilk olarak ikili sınıflandırma için geliştirilmiştir. Ancak, birçok görüntü tanıma uygulaması ikiden fazla sınıfa gereksinim duyar. Çoklu sınıf DVM problemleri, birçok ikili sınıflandırıcı oluşturularak çözülür (Yavuz ve Çevikalp 2008). One Against One (OAO)

ve One Against All (OAA) çoklu sınıf probleminin çözümü için yaygın olarak kullanılan iki yöntemdir. Her sınıf için OAA, söz konusu sınıfın üyelerini diğer sınıfların üyelerinden ayırmak için bir adet ikili DVM kullanır. Diğer bir yandan, OAO ise her sınıf için bir tane ikili DVM kullanmak yerine söz konusu sınıfın üyelerini diğer sınıfların üyelerinden ayırmak için her bir sınıf çifti için bir tane ikili DVM kullanır. OAO, OAA'dan biraz daha yavaş olmasına rağmen daha doğru sonuçlar üretir (Aisen 2006). Bu çözüm, sınıflandırma işlemi karmaşık hale getirir. RO sınıflandırıcısı DVM'nin bu dezavantajına bir alternatif oluşturmaktadır. Bu anlamda RO, çok sayıda değişken ve sınıfı, karmaşık modellemelere ve parametrelere ihtiyaç duymadan sınıflandırabildiği için DVM yöntemine bir alternatif oluşturmaktadır. Ayrıca RO, görüntüye karşı hassasiyet göstermediği ve modele uydurma söz konusu olmadığı için birçok ağaç tabanlı algoritmadan da üstündür (Watts ve Lawrence 2008).

Literatürde RO sınıflandırıcısının performansını diğer sınıflandırıcılardan elde edilen sonuçlarla karşılaştırarak test eden birçok çalışma vardır. Jayvd. (2009), RO yönteminin hem karmaşık hem de homojen yeşil alan topluluklarını %88.37 genel sınıflandırma doğruluğuyla başarılı bir şekilde sınıflandırdığını belirtmektedir. Waske ve Braun (2009) öğrenme tabanlı (RO ve Hızlandırma) ve EÇB yöntemlerini kullanarak zamansal SAR görüntülerini sınıflandırmışlardır. Bu çalışma öğrenme tabanlı yöntemlerin EÇB'den yaklaşık %10 daha yüksek sınıflandırma doğruluğu verdiğini göstermektedir. Prasad vd. (2006), Regression Tree Analysis (RTA), RO, Torbalama Ağaçları (TA) ve Multivariate Adaptive Regression Splines (MARS) yöntemlerini kullanarak dört tür için yeşil alan örtüsü haritaları oluşturmuşlardır. Kappa ve türevleri, değişken önemi ve çıktı haritaları gibi birden çok istatistiksel değerlendirme ölçütlerine ait çıktıları değerlendirerek bu dört metodu karşılaştırmıştır. Sonuçlar TA ve RO'nun diğerlerinden üstün olduğunu, RO'nun az da olsa daha iyi performans sergilediğini göstermiştir. Ayrıca Watts ve Lawrence (2008), Waske vd. (2007), Gislason vd. (2004), Pal (2003) ve Akar vd. (2010) çalışmalarında RO'nun hızlı olduğunu ve yüksek doğruluk gösterdiğini vurgulamışlardır.

Bu çalışma, oylama tabanlı toplu öğrenme sınıflandırma yöntemi olarak bilinen Rastgele Orman (RO) algoritmasının performansını incelemektedir. RO'dan elde edilen sonuçlar, kırsal ve kentsel alanlara ait farklı çözünürlükteki uydu görüntüleri kullanılarak Gentle AdaBoost (GAB), EÇB ve DVM algoritmalarından elde edilen sonuçlarla karşılaştırılmıştır. Hem toplu öğrenme yöntemi hem de makine öğrenme yöntemi olan RO, toplu öğrenme yöntemi olan GAB ve makine öğrenme yöntemi olan DVM ile karşılaştırılmıştır. RO, obje tabanlı yaklaşımlar gibi pikselleri bir grup olarak değil bireysel olarak ele aldığı için piksel tabanlı sınıflandırma algoritmaları arasında da sayılabilir. Bu nedenle, RO klasik piksel tabanlı sınıflandırma algoritması olarak yaygın olarak kullanılan EÇB yöntemiyle de karşılaştırılmıştır.

2. Rastgele orman

Toplu sınıflandırma yöntemleri, bir sınıflandırıcı yerine birden çok sınıflandırıcı üreten ve sonrasında onların tahminlerinden alınan oylar ile yeni veriyi sınıflandıran öğrenme algoritmalarıdır. En yaygın olarak kullanılan toplu sınıflandırıcılar, Torbalama, Hızlandırma ve RO'dur.

Torbalama algoritmasında, bir sınıflandırıcıyı eğitmek için yer değiştirmeli olarak orijinal eğitim veri setinden birden çok

önyüklemeli eğitim veri setleri oluşturulur ve her bir önyüklemeli eğitim veri seti için bir ağaç üretilir. Ardışık gelen ağaçlar bir öncekinden bağımsızdır ve tahmin için en büyük oy baz alınır (Liaw ve Wiener 2002). Diğer taraftan, Hızlandırma iteratif olarak tekrarlı eğitimi kullanır ve yanlış sınıflandırılmış örnekler bundan sonraki iterasyonda daha önemli olsun diye bunların ağırlıkları artırılır. Hızlandırma genel olarak sınıflandırmanın hem varyansını hem de sapmasını azaltır ve çoğu durumda Torbalama'dan daha doğru sonuç verir. Ancak bazı dezavantajlara sahiptir. Yavaşdır, tekrarlı eğitim olabilir ve gürtlüye karşı hassastır (Gislason vd. 2006).

RO, ağaç tipi sınıflandırıcılar topluluğu olarak tanımlanabilir. Rastgelelik özelliği eklenerek Torbalama yönteminin geliştirilmiş bir versiyonudur (Breiman 2001). RO, tüm değişkenler arasından en iyi dalı kullanarak her bir düğümü dallara ayırmak yerine, her bir düğümde rastgele olarak seçilen değişkenler arasından en iyisini kullanarak her bir düğümü dallara ayırır. Her bir veri seti orijinal veri setinden yer değiştirmeli olarak üretilir. Sonra rastgele özellik seçimi kullanılarak ağaçlar geliştirilir. Geliştirilen ağaçlar budanmaz (Archer 2008; Breiman 2001). Bu strateji RO'nun doğruluğunu eşsiz yapar (Breiman ve Cutler 2005). RO aynı zamanda çok hızlıdır, aşırı uyuma karşı dayanıklıdır ve ne kadar istenirse o kadar ağaçla çalışılır (Breiman ve Cutler 2005).

RO algoritmasını başlatmak için kullanıcı tarafından 2 parametre tanımlanmalıdır. Bu parametreler, en iyi bölünmeyi belirlemek için her bir düğümde kullanılan değişkenlerin sayısı (m) ve geliştirilecek ağaçların sayısı N 'dir. İlk olarak eğitim veri setinin $2/3$ 'ünden önyükleme örnekleri oluşturulur. Out-of-bag (OOB) verisi olarak da adlandırılan, eğitim veri setinin $1/3$ 'lük geri kalan kısmı hataları test etmek için kullanılır. Sonra her bir önyüklemeli örnekten budama olmadan ağaç geliştirilir. Her bir düğümde m değişkenleri tüm değişkenler arasından rastgele olarak seçilir ve bu değişkenler arasından en iyi dal belirlenir. Yeterli öngörü gücü ile yeterli miktarda düşük korelasyon sağlayan değişken sayısının seçimi son derece önemlidir (Horning 2010). Breiman (2002)'e göre toplam M adet değişken sayısının kare köküne eşit alınan m değişken sayısı genel olarak optimum sonuca en yakın sonucu verir. RO ağaç üretmek için CART (Classification and Regression Tree) algoritmasını kullanır (Breiman 2001). Her bir düğümde dallar CART algoritmasının kriterine (örn. GINI indeksi) göre oluşturulur. GINI indeksi sınıf homojenliğini ölçer ve aşağıdaki formül (1) ile ifade edilebilir.

$$\sum \sum_{j \neq i} (f(C_i, T) / |T|) (f(C_j, T) / |T|) \quad (1)$$

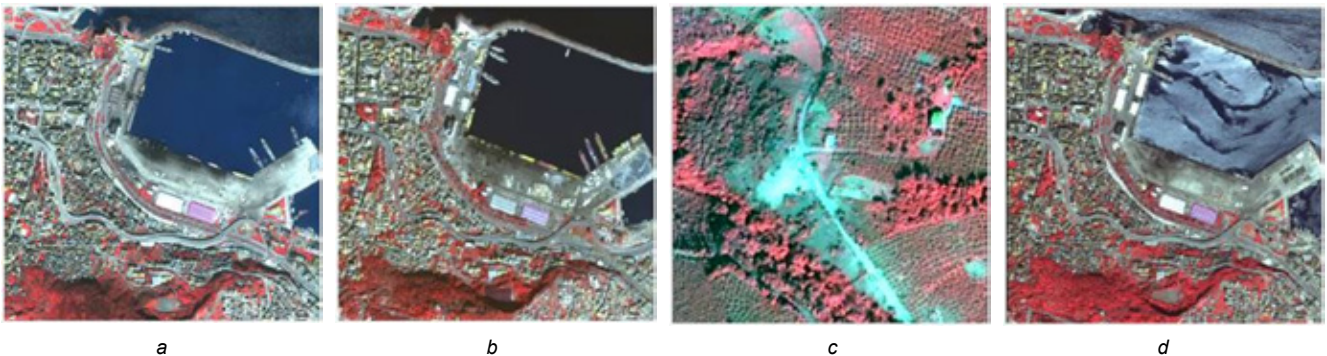
Burada T eğitim veri setini, C_i rastgele seçilen bir pikselin ait

olduğu sınıfı C_i ve $f(C_i, T) / |T|$, seçilen örneğin C_i sınıfına ait olma olasılığını göstermektedir (Pal 2005). GINI indeksi büyüdükçe sınıf heterojenliği artarken, GINI indeksi azaldıkça sınıf homojenliği artar. Bir alt düğümün GINI indeksi bir üst düğümün GINI indeksinden daha az olduğunda o dal başarılıdır. GINI indeksi sıfıra ulaşıncaya yani her bir yaprak düğümde bir sınıf kaldığında ağaç dallanma işlemi sonlanır (Watts vd. 2011). Tüm N adet ağaç üretildiğinde N tane ağaçtan elde edilen tahmin sonuçlarına dayanılarak aday pikselin sınıfı belirlenir (Liaw ve Wiener 2002).

Yukarıda açıklanan RO yöntemi ile görüntü sınıflandırma işlemi şu şekilde gerçekleştirilir. Farz edelim ki toplam ağaç sayısı (N) 1000 olarak seçilmiş olsun. Algoritma, tek bir piksel için 1000 farklı ağaç, yani 1000 farklı sınıflandırma sonucu üretir. Bunların içerisinde 800 ağaçta söz konusu piksel orman olarak sınıflandırılmış olsun. Aynı piksel 100 ağaçta arazi, kalan 100 ağaçta ise su olarak sınıflandırılmış olsun. Bu piksel için belirlenen sınıf orman olacaktır.

3. Çalışma alanı ve veri

Bu çalışma, Türkiye'nin Trabzon ilinde belirlenen kırsal ve kentsel alanlara ait birden çok uydu görüntüsü kullanılarak yapılmıştır. Kullanılan veriler 2005 ve 2008 yıllarında alınan QuickBird Pansharpened çok bantlı görüntüleri (0.6 m) ve 2003 yılında alınan Ikonos çok bantlı görüntüsünü (4 m) içerir (Şekil 1). Eğitim veri seti görüntü üzerinden her bir sınıf için görsel olarak birden fazla poligon belirlenmesi ve sayısallaştırılması ile seçilmiştir. Kentsel alan için görüntü üzerinden Deniz, Yeşil Alan, Toprak, Kentsel Yapı 1, Kentsel Yapı 2, Kentsel Yapı 3, Petrol Kalıntısı ve Gölge olmak üzere 8 farklı sınıf için eğitim alanları seçilmiştir. Kırsal bölge için ise Orman 1, Orman 2, Meyve Bahçesi, Çimen, Toprak, Yol, Bina ve Gölge gibi farklı arazi kullanım sınıfları kullanılmıştır. Sonrasında her bir görüntü üzerinde seçilen bu örnek alanlara göre eğitim ve test verileri Rastgele Özellik Seçim Metodu kullanılarak Matlab programında oluşturulmuştur. Bu eğitim alanları daha sonra RO ve GAB yöntemleri ile görüntü sınıflandırma işleminde kullanılmıştır. IKONOS, Quickbird (2005), Quickbird (2008)-Kentsel, ve Quickbird (2008)-Kırsal görüntüleri için eğitim veri setlerinin toplam sayısı sırasıyla 6381, 15092, 12021 ve 7716 pikseldir. Her sınıf için yaklaşık olarak eşit sayıda eğitim için, geriye kalan kısım da test verisi için kullanılmıştır. DVM ve EÇB için de yine aynı eğitim alanları kullanılmıştır. RO ve GAB için Matlab kodu, DVM ve EÇB için ise ENVI yazılımı kullanılarak görüntüler sınıflandırılmıştır.



Şekil 1: Çalışma Alanları a) Kentsel Alana ait Ikonos çok bantlı görüntüsü, b) Kentsel alana ait QuickBird pansharpened görüntüsü (2008 tarihli), c) Kırsal alana ait QuickBird pansharpened görüntüsü (2008 tarihli), d) Gürültü efekti içeren kentsel alana ait QuickBird pansharpened görüntüsü (2005 tarihli).

4. Sonuçlar ve irdelemeler

RO yönteminin sınıflandırma doğruluğu, N (ağaç sayısı) ve m (parametre sayısı) gibi kullanıcı tanımlı parametrelere bağlıdır. Bundan dolayı veri için en uygun parametrenin seçimi sınıflandırma doğruluğunu artırır. Çalışmada, N ve m için optimum parametreleri bulabilmek için çoklu kombinasyonlar test edilmiş ve tematik haritalarının daha doğru olarak elde edilebilmesi için değerlendirilmiştir (Tablo 1). Eğitim veri seti kullanılarak farklı N ve m kombinasyonları için elde edilen OOB hatası, test doğruluğu, kapa ve hesaplama zamanı değerleri Tablo 1’de verilmiştir. Tablo 1’de görüldüğü gibi, kentsel alana ait Ikonos görüntüsü için $N = 100$ ve $m = 2$ seçilmiştir. Kırsal alana ait QuickBird görüntüsü için $N = 500$ ve $m = 2$ seçilmesine karşın; kentsel alana ait QuickBird görüntüsü için $N = 350$ ve $m = 2$ seçilmiştir.

Her bir görüntü için optimum parametreler belirlendikten sonra görüntüler RO algoritması ile sınıflandırılmıştır. Sınıflandırma işlemleri sonucu elde edilen tematik haritalar Şekil 2 ve Şekil 3’te gösterilmiştir. Her bir sınıflandırma sonucunun doğruluğu, en yaygın kullanılan sınıflandırma sonrası doğruluk analizleri yöntemlerinden biri olan hata matrisi kullanılarak değerlendirilmiştir. Bu matris kullanılarak, bilinen referans veriler (yer gerçeği) ile buna karşılık gelen otomatik sınıflandırma sonucu elde edilen sonuçlar arasındaki ilişkiler kategorik olarak birbirleriyle karşılaştırılabilmektedir. (Lillesand vd. 2004).

Doğruluk analizi için toplamda 240 adet nokta görüntü üzerine rastgele dağıtılmıştır. Her sınıf için atılan rastgele nokta sayısı, o sınıfın görüntü üzerinde kapladığı alanla orantılı olacak şekilde belirlenmiştir. Her bir tematik haritanın doğruluğu aynı 240 nokta kullanılarak test edilmiştir. Tablo 3, kentsel ve kırsal alan için RO’nun hata matrislerini göstermektedir.

RO algoritmasının performansını değerlendirmek için RO’nun sınıflandırma sonuçları GAB, DVM ve EÇB algoritmalarından elde edilen sonuçlarla karşılaştırılmıştır. RO algoritmasına benzer şekilde DVM de en uygun parametre seçimi gerektirir. DVM’nin ENVI yazılımında uygulanması esnasında, çoklu sınıf problemleri için OAO olarak ta bilinen çiftli sınıflandırma stratejisi kullanılır. DVM sınıflandırma işleminde, en yüksek doğruluğu sağlayan kernel fonksiyonlarından biri olan Radyal Tabanlı Fonksiyon kerneli (Kavzoğlu ve Çolkesen 2010) kullanılmıştır. Her bir görüntü için en uygun C parametresi belirlenmiş ve farklı C parametre kombinasyonları Tablo 2’de verilmiştir.

Tüm yöntemlerin hata matrisleri analiz edildiğinde, benzer spektral özelliğe sahip olan sınıfların büyük ihtimalle birbirleriyle karıştığı görülmektedir. Kentsel alana ait Ikonos görüntüsünde Kentsel Yapı 2, Kentsel Yapı 3, Yeşil Alan

ve Toprak sınıfları benzer spektral özelliğe sahip oldukları için en çok karışan sınıflar olmuşlardır. Petrol Kalıntısı ve Gölge sınıfları da benzer spektral özelliklere sahiptir. Tablo 4’te görüldüğü üzere özellikle Yeşil Alan sınıfı için RO’nun, GAB, DVM ve EÇB’den yaklaşık %22 daha yüksek üretici doğruluğuna sahip olduğu sonucu çıkarılabilir. Aynı sınıfın kullanıcı doğruluğu incelendiğinde RO, GAB’den yaklaşık %28, DVM’den yaklaşık %12 ve EÇB’den yaklaşık %4 daha iyi performans sergilediği görülmektedir. Kentsel Yapı 3 sınıfı için DVM, RO’dan %1 daha yüksek performans sergilediği için en iyi performansa sahip olmuştur. Toprak sınıfı için RO, SVM (~%19) ve EÇB(~%13)’den daha yüksek üretici doğruluğuna sahiptir. Aynı sınıfın kullanıcı doğruluğu EÇB’nin en düşük performansa (RO’dan %43 daha düşük) sahip olduğunu göstermektedir.

Ikonos görüntüsünden daha yüksek bir konumsal çözünürlüğe sahip olan QuickBird-2005 görüntüsünde RO en yüksek sınıflandırma doğruluğunu (%83.75) vermiştir. Hem kullanıcı hem de üretici doğrulukları incelendiğinde, Tablo 4 RO’nun Yeşil Alan, Toprak, Kentsel Yapı 2, Kentsel Yapı 3 ve Gölge sınıflarını diğer yöntemlerden daha iyi ayırdığını ortaya koymaktadır. Üretici doğruluklarına göre RO, Toprak, Kentsel Yapı 3, Gölge, Petrol Kalıntısı ve Yeşil Alan sınıflarında diğer yöntemlerden daha iyi doğruluğa sahiptir.

Kırsal için sınıflandırma sonuçları, RO algoritmasının GAB, DVM ve EÇB’den daha başarılı olduğunu göstermektedir (Tablo 5). Orman, Çimen ve Meyve Bahçesi sınıfları benzer spektral özelliğe sahiptir. Orman 1 sınıfı için RO, GAB’den (yaklaşık olarak % 29), DVM’den (yaklaşık olarak % 21) ve EÇB’den (yaklaşık olarak % 38) daha yüksek üretici doğruluğuna sahiptir. Bunun yanında aynı sınıf için RO, DVM’nin kullanıcı doğruluğunu yaklaşık % 17 ve EÇB’nin kullanıcı doğruluğunu ise yaklaşık % 29 artırmıştır. Çimen sınıfında da benzer sonuçlar görülmüştür. Meyve Bahçesi sınıfı için EÇB üretici doğruluğunda en iyi performansı sergilerken, RO kullanıcı doğruluğunda en iyi doğruluğa sahip olmuştur.

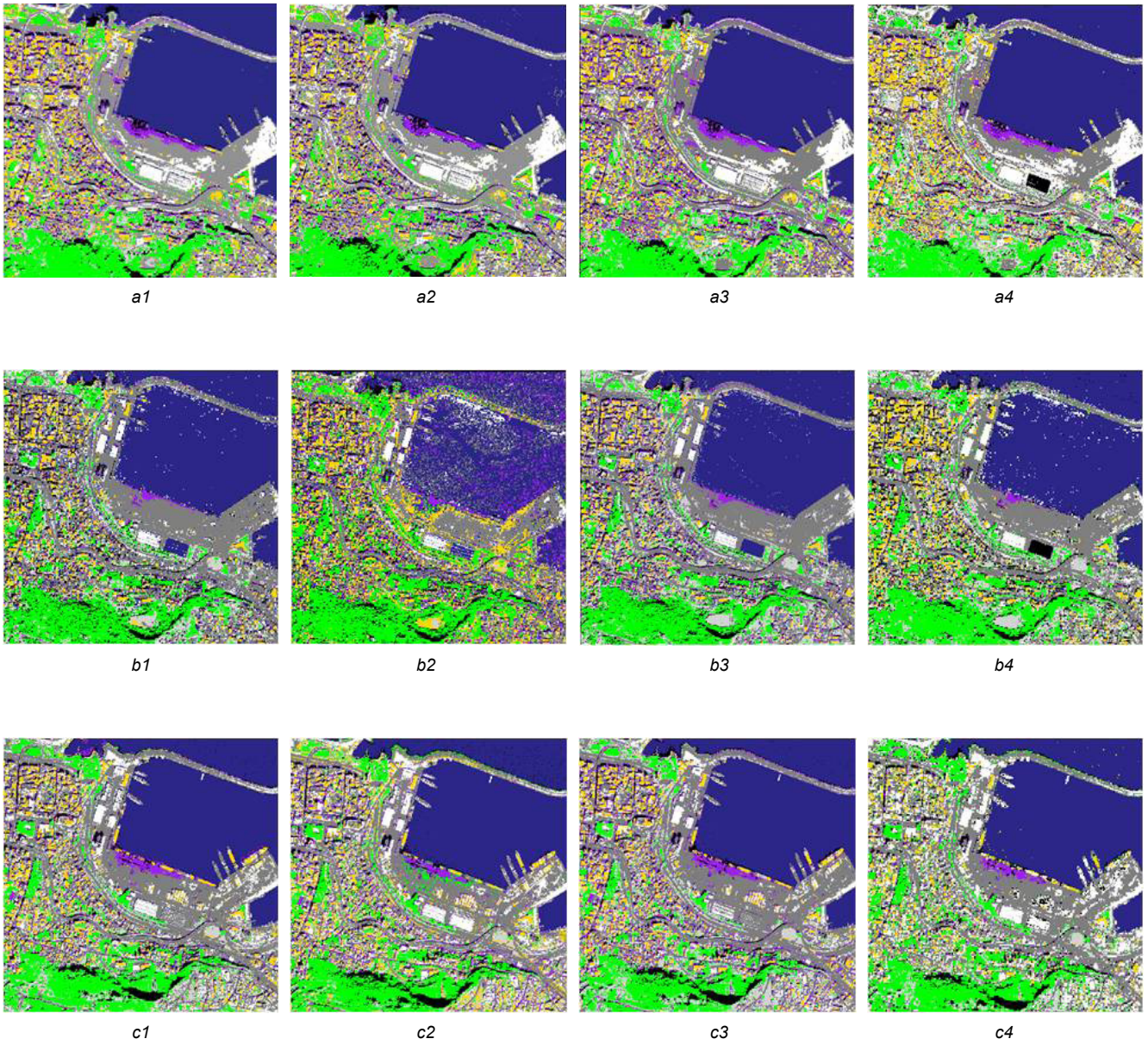
QuickBird 2005 görüntüsünün seçilmesinin nedeni, RO algoritmasının performansının gürültü efekti olan bir görüntü kullanılarak test edilmesidir. Bu görüntü rüzgarlı bir günde alındığından dolayı Karadeniz üzerindeki bölümler homojen değildir. Dalgalar sebebiyle pürüzlü olan deniz yüzeyi, makalede gürültü olarak nitelendirilen gölgeleri ve beyaz köpüklerin oluşmasına (sanki görüntüye bilerek gürültü ilave edilmiş gibi) sebep olmuştur. Sonuç olarak, deniz yüzeyindeki bu gürültü örüntülerine ait pikseller deniz yerine Kentsel Yapı 1 sınıfı olarak sınıflandırılmıştır. Tablo 7 ve Tablo 8’de görüldüğü gibi, deniz yüzeyindeki gürültü örüntüleri sınıflandırıcıların performansını olumsuz yönde etkilemiştir. Ancak, RO algoritması bu istenmeyen durumdan en az etkilenmiş ve %83.75’lik genel

Tablo 1: Ikonos görüntüsü için RO parametrelerinin testi

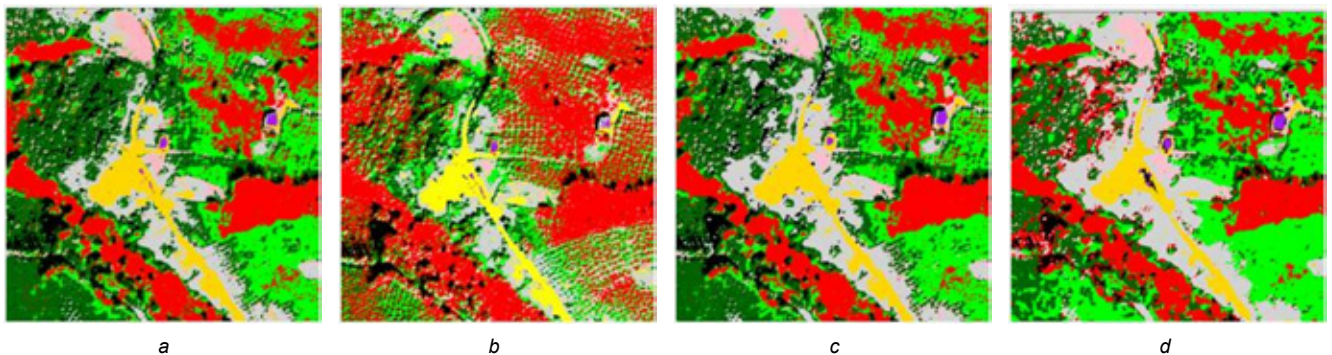
N	m	OOB Hatası (%)	Doğruluk (%)	Kappa	Hesaplama Zamanı (sn)
95	2	0.0536	78.33	0.7521	34.2965
100	2	0.0465	87.08	0.8524	36.2796
125	2	0.0444	80.42	0.7757	42.2122
233	2	0.0486	78.75	0.7563	69.9202
250	2	0.0465	77.92	0.7477	78.1638
500	2	0.0514	81.67	0.7891	144.4255

Tablo 2: Kırsal alana ait 2008 QuickBird görüntüsü için DVM yönteminin farklı C parametrelerinin test sonuçları

C	Test Doğruluğu (%)	Kappa
50	66.40	0.5836
60	66.80	0.5884
70	67.20	0.5937
80	68.00	0.6038
90	68.00	0.6044
100	67.60	0.5992



Şekil2:Kentsel alana ait sınıflandırma sonuçları. İkonos için sınıflandırma sonuçları: RO(a1), GAB (a2), DVM (a3) ve EÇB (a4). Gürültü efekti olan QuickBird-2005 için sınıflandırma sonuçları: RO(b1), GAB (b2), DVM (b3) ve EÇB (b4). QuickBird-2008 için sınıflandırma sonuçları: RO(c1), GAB (c2), DVM (c3) ve EÇB (c4).



Şekil3: Kırsal alana ait QuickBird-2008görüntüsünün sınıflandırma sonuçları: RO (a), DVM (b), GAB (c)ve EÇB (d) yöntemleri.

Tablo 3: RO'nun hata matrisi a) Kentsel Alan (Ikonos) b) Kırsal Alan (QuickBird -2008)

		Sınıflar										Üretici Doğruluk(%)	Kullanıcı Doğruluk(%)
		Deniz	Bitki	Toprak	Kentsel Yapı 1	Kentsel Yapı 2	Petrol Kalıntısı	Kentsel Yapı 3	Gölge	Satır Toplamı			
Sınıflar	Deniz	30	0	0	0	0	0	0	0	30	100.00%	100.00%	
	Bitki	0	30	0	0	0	0	0	0	30	93.75%	100.00%	
	Toprak	0	0	30	0	0	0	0	0	30	96.77%	100.00%	
	Kentsel Yapı 1	0	0	0	30	0	0	0	0	30	93.75%	100.00%	
	Kentsel Yapı 2	0	1	0	1	27	0	0	1	30	77.14%	90.00%	
	Petrol Kalıntısı	0	0	1	0	7	7	2	13	30	87.50%	23.33%	
	Kentsel Yapı 3	0	0	0	1	1	0	27	1	30	93.10%	90.00%	
	Gölge	0	1	0	0	0	1	0	28	30	65.12%	93.33%	
Sütun Toplamı	30	32	31	32	35	8	29	43	240				

O Genel Doğruluk= 87.08% Kappa= 0.8524

a)

		Sınıflar										Üretici Doğruluk(%)	Kullanıcı Doğruluk(%)
		Orman 1	Orman 2	Meyve Bahçesi	Çimen	Toprak	Yol	Kentsel Yapı	Gölge	Satır Toplamı			
Sınıflar	Orman 1	21	0	5	1	0	0	0	3	30	87.50%	70.00%	
	Orman 2	0	19	8	0	0	0	0	3	30	70.37%	63.33%	
	Meyve Bahçesi	0	7	22	1	0	0	0	0	30	62.86%	73.33%	
	Çimen	3	1	0	26	0	0	0	0	30	89.66%	86.67%	
	Toprak	0	0	0	0	24	5	0	1	30	92.31%	80.00%	
	Yol	0	0	0	1	2	27	0	0	30	84.38%	90.00%	
	Kentsel Yapı	0	0	0	0	0	0	30	0	30	100.00%	100.00%	
	Gölge	0	0	0	0	0	0	0	30	30	81.08%	100.00%	
Sütun Toplamı	24	27	35	29	26	32	30	37	240				

Genel Doğruluk= 82.92% Kappa= 0.8048

b)

Tablo 4: Görüntülerin kentsel alan için kullanıcı ve üretici doğrulukları

Sınıflar	Ikonos								QuickBird-2008							
	RO		GAB		DVM		EÇB		RO		GAB		DVM		EÇB	
	Üretici Doğruluk(%)	Kullanıcı Doğruluk(%)	Üretici Doğruluk(%)	Kullanıcı Doğruluk(%)	Üretici Doğruluk(%)	Kullanıcı Doğruluk(%)	Üretici Doğruluk(%)	Kullanıcı Doğruluk(%)	Üretici Doğruluk(%)	Kullanıcı Doğruluk(%)	Üretici Doğruluk(%)	Kullanıcı Doğruluk(%)	Üretici Doğruluk(%)	Kullanıcı Doğruluk(%)	Üretici Doğruluk(%)	Kullanıcı Doğruluk(%)
Deniz	100.00	100.00	100.00	83.33	96.67	96.67	100.00	90.91	93.75	100.00	96.88	93.94	96.88	96.88	93.75	100.00
Yeşil Alan	93.75	100.00	71.88	71.88	71.88	88.46	71.88	95.83	81.08	100.00	70.27	61.90	62.16	95.83	81.08	81.08
Soil	96.77	100.00	38.71	80.00	77.42	77.42	83.87	56.52	81.48	73.33	48.15	61.90	74.07	54.05	62.96	68.00
Kentsel Yapı 1	93.75	100.00	90.63	93.55	87.50	93.33	84.38	87.10	93.33	93.33	93.33	93.33	80.00	96.00	96.67	63.04
Kentsel Yapı 2	77.14	90.00	85.71	54.55	80.00	59.57	74.29	74.29	90.00	90.00	80.00	82.76	90.00	81.82	76.67	56.10
Petrol Kalıntısı	87.50	23.33	62.50	20.83	75.00	18.75	87.50	50.00	100.00	36.67	72.73	33.33	81.82	29.03	72.73	57.14
Kentsel Yapı 3	93.10	90.00	62.07	81.82	72.41	91.30	72.41	65.63	96.00	80.00	64.00	55.17	92.00	82.14	52.00	76.47
Gölge	65.12	93.33	48.84	84.00	44.19	90.48	41.86	94.74	62.50	100.00	62.50	93.75	58.33	93.33	54.17	92.86

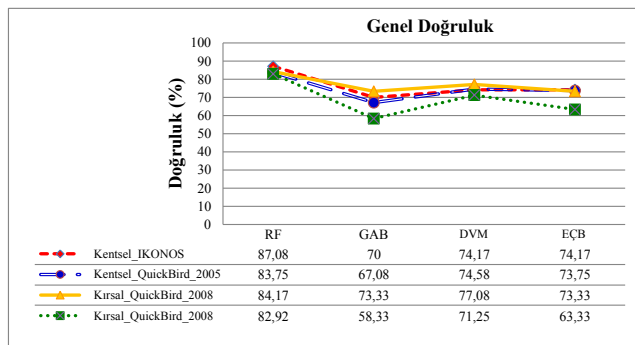
Tablo 5: Kırsal alana ait QuickBird-2008 görüntüsü için kullanıcı ve üretici doğrulukları

	RO		GAB		DVM		EÇB	
	Üretici Doğruluğu(%)	Kullanıcı Doğruluğu(%)	Üretici Doğruluğu(%)	Kullanıcı Doğruluğu(%)	Üretici Doğruluğu(%)	Kullanıcı Doğruluğu(%)	Üretici Doğruluğu(%)	Kullanıcı Doğruluğu(%)
Orman 1	87.50	70.00	58.33	1.62	66.67	53.33	50.00	41.38
Orman 2	70.37	63.33	85.19	41.82	70.37	57.58	55.56	38.46
Meyve Bahçesi	62.86	73.33	14.29	50.00	57.14	68.97	71.43	53.19
Çimen	89.66	86.67	75.86	68.75	89.66	83.87	55.17	80.00
Toprak	92.31	80.00	69.23	48.65	75.00	63.16	78.13	100.00
Yol	84.38	90.00	75.00	68.57	88.46	63.89	100.00	65.00
Kentsel Yapı	100.00	100.00	30.00	100.00	60.00	100.00	76.67	100.00
Gölge	81.08	100.00	67.57	96.15	67.57	100.00	27.03	100.00

Tablo 6: Gürültü efekti olan QuickBird-2005 görüntüsü için kullanıcı ve üretici doğrulukları

	RO		GAB		DVM		EÇB	
	Üretici Doğruluğu(%)	Kullanıcı Doğruluğu(%)	Üretici Doğruluğu(%)	Kullanıcı Doğruluğu(%)	Üretici Doğruluğu(%)	Kullanıcı Doğruluğu(%)	Üretici Doğruluğu(%)	Kullanıcı Doğruluğu(%)
Deniz	90.63	96.67	84.38	79.41	90.63	96.67	87.50	96.55
Yeşil Alan	85.71	100.00	94.29	73.33	85.71	85.71	85.71	78.95
Toprak	95.00	63.33	25.00	41.67	95.00	41.30	85.00	53.13
Kentsel Yapı 1	96.55	93.33	75.86	91.67	79.31	95.83	65.52	76.00
Kentsel Yapı 2	80.56	96.67	41.67	65.22	63.89	82.14	69.44	67.57
Petrol Kalıntısı	100.00	23.33	100.00	20.59	100.00	25.00	100.00	50.00
Kentsel Yapı 3	96.67	96.67	96.67	65.91	73.33	100.00	80.00	70.59
Gölge	58.82	100.00	45.10	95.83	50.98	96.30	52.94	96.43

Tablo 7: RO, DVM, GAB ve EÇB yöntemlerinin kentsel ve kırsal alandaki genel doğrulukları



Tablo 8: Gürültü efekti olan/ gürültü efekti olmayan kırsal ve kentsel alan için ortalama genel doğruluklar

Gürültü Efekti	Alan	RO (%)	GAB(%)	DVM (%)	EÇB(%)
Yok	Kentsel	85.63	71.67	75.63	73.75
	Kırsal	83.75	67.08	74.58	73.75
Var	Kentsel	85.42	68.54	74.38	73.96

doğrulukla diğer sınıflandırıcılar arasında en iyi sınıflandırma performansını sergilemiştir. Bununla birlikte GAB, %68.54'lik sınıflandırma doğruluğuyla en düşük performansı göstermiştir. Bunun sonucu olarak, GAB yönteminin gürültüden en çok etkilendiği söylenebilir. Ayrıca, gürültüden büyük oranda etkilenen Kentsel Alan 1 ve Deniz sınıflarında RO ve DVM yöntemleri GAB ve EÇB'den daha iyi performans göstermiştir (Tablo 6).

5.Sonuç

Bu çalışma, farklı özellikteki bölgeler ve farklı çözünürlükteki görüntüler kullanarak RO algoritmasının performansını incelemektedir. RO algoritmasının sonuçları DVM, GAB ve EÇB yöntemlerinden elde edilen sonuçlarla karşılaştırılmıştır. İki tip görüntü kullanılmıştır; biri kentsel alan içeren İkonos ve QuickBird görüntülerinden oluşmaktadır. Diğer görüntü ise kırsal alana ait QuickBird görüntüsüdür. Ek olarak, gürültülü bir yapı içeren QuickBird görüntüsü ile görüntüdeki gürültünün sınıflandırma doğruluğuna etkisi araştırılmıştır. Sonuçlar, RO'nun %85.63 doğrulukla en yüksek doğruluğu verdiğini göstermektedir. Ayrıca RO, kentsel ve kırsal alan için DVM sonuçlarıyla karşılaştırıldığında sırasıyla %10 ve %15 daha iyi performans göstermiştir. Bu sonuçlar DVM'nin dört sınıflandırıcı arasında kırsal ve kentsel alanda en iyi ikinci sonucu veren sınıflandırıcı olduğunu ortaya koymaktadır. Gürültü efekti olan kentsel alan verisinde RO, DVM ve EÇB yöntemlerinin doğruluklarını yaklaşık %11 ve GAB'nin doğruluğunu yaklaşık %17 artırmıştır ki bu da gürültü efekti olan görüntüde RO'nun başarısını göstermektedir. Aynı zamanda RO'nun benzer spektral özellikteki sınıfları ayırmada başarılı olduğu söylenebilir. Bundan sonraki çalışmalar, RO'nun uydu görüntülerini sınıflandırmasındaki performansının geliştirilmesi için doku, eğitim ve diğer spektral olmayan bilgilerin RO yöntemine entegrasyonu üzerine olacaktır.

Kaynaklar

- Aisen B., (2006), *A comparison of multi class SVM methods*, <http://courses.media.mit.edu/2006fall/mas622j/Projects/aisen-project/index.html> [Erişim 20 Eylül 2012].
- Akar Ö., Güngör O., Akar A., (2010), *Rastgele orman sınıflandırıcısı ile arazi kullanım alanlarının belirlenmesi*, III. Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu, 11-13 Ekim 2010, Gebze, Kocaeli, Bildiriler Kitabı, 142-152.
- Archer K.J., (2008). *Empirical characterization of random forest variable importance measure, computational statistical data analysis*, Computational Statistics & Data Analysis, 52(4), 2249-2260.
- Breiman L., Cutler A., (2005), *Random forest*, http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm, [Erişim 12 Temmuz 2011].
- Breiman L., (2002), *Manual on setting up, using, and understanding random forests V3.1*, http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf, [Erişim 20 Eylül 2011].
- Breiman L., (2001), *Random forests, machine learning*, 2001 Kluwer Academic Publishers, 45(1), 5-32.
- Dronova I., Gong P., Wang L., (2011), *Object-based analysis and change detection of major wetland cover types and their classification uncertainty during the low water period at Poyang Lake, China*, Remote Sensing of Environment, 115(12), 3220-3236.

- ENVI, (2005), ENVI's User Guide, http://geol.hu/data/online_help/ApplyingMaximumLikelihoodClassification.html, [Erişim 20 Eylül 2011].
- Gao J.,(2009), *Digital analysis of remotely sensed imagery*, The McGraw-Hill Companies, USA.
- Gislason P.O.,Benediktsson J.A., Sveinsson J.R., (2006), *Random forest for land cover classification*, Pattern Recognition Letters, 27, 294-300.
- Gislason P.O.,Benediktsson J.A., Sveinsson J.R., (2004), *Random forest classification of multi-source remote sensing and geographic data*, IEEE International Geoscience and Remote Sensing Symposium IGARSS '04 Proceedings, 2, 1049 – 1052.
- Horning N., (2010), *Random Forests : An algorithm for image classification and generation of continuous fields data sets*, International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences (GIS-IDEAS) 2010, 9-11 December, Hanoi, Vietnam, wgrass.media.osaka-cu.ac.jp/gisideas10/viewpaper.php?id=342, [Erişim 15 Eylül 2011].
- Jay S., Lawrence R., Repasky K., Keith C., (2009), *Invasive species mapping using low cost hyper spectral imagery*, ASPRS 2009 Annual Conference, March 9-13, Baltimore, Maryland.
- Kaban Z., Diri B., (2008), *Genre and author detection in turkish texts using artificial immunere cognition systems*, IEEE 16th Signal Processing, Communication and Applications Conference, 20-22 April 2008, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4632548>, [Erişim 06 Ekim 2011].
- Kavzoğlu T.,Çölkesen İ., (2010), *Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi*, Harita Dergisi, 144.
- Liaw A.,Wiener M., (2002), *Classification and regression by random forest*, R News, 2(3).
- Lillesand T. M.,Kiefer R. W. And Chipman J. W., (2004), *Remote Sensing and Image Interpretation*, Wiley, United States of America, 804 ss.
- Matinfar H.R.,Sarmadian F., AlaviPanah S.K., Heck R.J., (2007), *Comparisons of object-oriented and pixel-based classification of land use/land cover types based on landsat7, etm+ spectral bands (case study: arid region of Iran)*, American-Eurasian J. Agric. & Environ. Sci., 2 (4), 448-456.
- Pal M., (2005), *Random forest classifier for remote sensing classification*, International Journal Of Remote Sensing, 26(1), 217-222.
- Pal M., (2003), *Random forest for land cover classification*, IEEE International Geoscience and Remote Sensing Symposium, IGARSS '03 Proceedings, 6, 3510-3512.
- Prasad A.M.,Iverson L.R.,Liaw A., (2006), *Newer classification and regression tree techniques: Bagging and random forests for ecological prediction*, Ecosystems, 9, 181-199.
- Waske B.,Braun M., (2009), *Classifier ensembles for land cover mapping using multi temporal SAR imagery*, ISPRS Journal of Photogrammetry and Remote Sensing, 64, 450-457.
- Waske B.,Heinzel V., Braun M., Menz G., (2007), *Random forests for classifying multi-temporal sar data*, Proc. 'Envisat Symposium 2007', 23-27 April 2007, Montreux, Switzerland, <http://envisat.esa.int/envisatsymposium/proceedings/sessions/3D3/461589wa.pdf>, [Erişim 11 Ağustos 2011].
- Watts J.D., Lawrence R.L., (2008), *Merging random forest classification with an object-oriented approach for analysis of agricultural lands*, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVII(B7).
- Watts J. D.,Powell S. L., Lawrence R. L., Hilker T., (2011), *Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery*, Remote Sensing of Environment 115, 66-75.
- Yavuz H. S, Çevikalp H., (2008), *A new distance measure for hierarchical clustering*, IEEE 16th Signal Processing, Communication and Applications Conference, 20-22 April 2008, Aydın, ss.1-4.