

RASTGELE ORMAN SINIFLANDIRMA YÖNTEMİ YARDIMIYLA TARIM ALANLARINDAKİ ÜRÜN ÇEŞİTLİLİĞİNİN SINIFLANDIRILMASI

A. Özdarıcı Ok^{a*}, Ö. Akar^{b*}, O. Güngör^b

^a ODTÜ, Jeodezi ve Coğrafi Bilgi Teknolojileri EABD, 06531 Çankaya, Ankara, Türkiye - ozdarici@metu.edu.tr

^b KTÜ, Harita Mühendisliği, 61080, Trabzon, Türkiye - ozlemerden@gmail.com

Commission VI, WG VI/4

ANAHTAR KELİMELER: Rastgele Orman Sınıflandırma Yöntemi, En Çok Benzelik Sınıflandırma Yöntemi, SPOT 5, Tarım, Doğruluk Analizi

ÖZET:

Çalışmada Rastgele Orman (RO) sınıflandırma yöntemi SPOT 5 görüntüsüne ait bir tarım alanı üzerinde uygulanmış ve elde edilen sonuçlar En Çok Benzerlik (EÇB) sınıflandırma yöntemi sonucu ile karşılaştırılarak RO sınıflandırma yönteminin ürün sınıflandırmasındaki başarısı incelenmiştir. Test alanı olarak zengin toprak yapısı ve ikliminin tarıma elverişli olması nedeniyle Bursa ilinde yer alan Karacabey Ovası (~100 km²) seçilmiştir. Test alanında yaklaşık 12 çeşit ürün yetiştirilmektedir. Yetiştirilen başlıca ürünler arasında mısır, domates, biber, pirinç, şeker pancarı ve buğday gelmektedir. Çalışmada SPOT5 uydu görüntüsünden 5 farklı ürün sınıfı (mısır, buğday, domates/biber, şeker pancarı ve pirinç) için örnek alanlar toplanmış ve bu örnek alanlar daha sonra RO sınıflandırma yöntemine girdi olarak kullanılmıştır. Aynı örnek alanlar daha sonra EÇB sınıflandırma yöntemi yardımıyla sınıflandırılarak elde edilen bulgular RO sınıflandırma yöntemi sonucunda üretilen sonuçlar ile karşılaştırılmıştır. Her iki yönteme ait sınıflandırma sonuçları alanda bulunan her parselde tek ürün yetiştirildiği varsayılarak parsel tabanlı olarak değerlendirilmiştir. Bu sayede sınıflandırma sonucunda üretilecek tematik haritanın doğruluğunu olumsuz yönde etkileyebilecek karışık piksel değerleri giderilmeye çalışılmıştır. Sonuçlar, ürün sınıfları için RO sınıflandırma yöntemi sonucunda elde edilen genel ortalama hata oranının %85.89, EÇB sınıflandırma yöntemi kullanılarak üretilen genel ortalama hata oranının %77.96 olduğunu göstermiştir. Buna göre RO sınıflandırma yönteminin EÇB sınıflandırma yöntemi sonucunu yaklaşık %8 oranında arttırdığı gözlenmiştir. Bu oran, RO sınıflandırma yönteminin tarım alanlarında yetiştirilen ürün çeşitliliğini tespit etmede önemli bir yöntem olabileceğini göstermektedir.

CLASSIFICATION OF CROPS IN AGRICULTURAL LANDS USING RANDOM FOREST CLASSIFICATION METHOD

ABSTRACT:

In this study, Random Forest (RF) image classification method is applied on an agricultural land using SPOT 5 satellite image and the performance of the RF classification is investigated by comparing the results obtained with the Maximum Likelihood Classification (MLC) method. Due to the characteristics of rich and loamy soils combined with reasonable weather conditions, Karacabey Plain (~100 km²) in Bursa was selected as a test site. Approximately 12 crop types are cultivated in the test area. The main crop types involve corn, tomato, pepper, rice, sugar beet and wheat. In the study, training areas are selected on SPOT 5 image for 5 different crop types (corn, wheat, tomato/pepper, sugar beet and rice) and these training areas are then utilized as an input to the RF classification. The same training areas are then used during the classification of the image with the MLC method and the classification results obtained are compared with the results of the RF classification method. The results of the image classification are evaluated in a parcel-based manner based on the assumption that each field contains only one crop type. In this way, the mixed pixel effects that may inversely affect the accuracy of the thematic map are tried to be reduced. Results indicate that the overall accuracies produced by the RF and MLC classification methods are 85.89% and 77.96%, respectively. Based on those results, it is observed that the RF classification method increased the accuracy of classification of the well-known MLC method around 8%. This rate indicates that the RF image classification method could be a reliable method to classify the crops cultivated in agricultural lands.

KEY WORDS: RF image Classification, MLC method, SPOT 5, Agriculture, Accuracy Assessment

* Corresponding authors. ODTÜ, Jeodezi ve Coğrafi Bilgi Teknolojileri EABD, 06531 Çankaya, Ankara, Türkiye
Email: [aозdarici@gmail.com](mailto:aozdarici@gmail.com) (A.Özdarıcı Ok)
KTÜ, Harita Mühendisliği, 61080, Trabzon, Türkiye
Email: ozlemerden@gmail.com (Ö.Akar)

1.GİRİŞ

Uydu görüntülerinin sınıflandırılması sonucu elde edilen tematik haritalar haritacılık, subilimi, ormancılık alanlarında ve zirai alanlarda yaygın olarak kullanılmaktadır. Örneğin, biyolojik çeşitlilik açısından büyük önem taşıyan ve zengin bir kaynak olan tropikal ormanların izlenmesi (Christian ve Krishnaya, 2009), arazi kullanım alanlarının belirlenmesi (Huang ve Fipps, 2006), su kaynaklarının yönetimi için su karakteristiklerinin (su, kuru, nemli içerikler...) tespiti (Villalon-Turrubiates, 2009), toprak nemliliği tahmini (Sajjad, 2010), kentsel ve çevresel değişim analizleri, nesne çıkarımları (Zhang, vd., 2007) tarımsal alanlarda toprak erozyonunu azaltmada önemli rol oynayan ürün kalıntılarının tespiti (Daughtry,2006) gibi birçok uygulama alanında sınıflandırılmış görüntülerden yararlanılmaktadır.

Bilgi çıkartmak olarak da adlandırılan görüntü sınıflandırma, görüntüdeki piksel değerlerinin kullanılarak bu bilgilerin, anlamlı arazi örtüsü bilgisine dönüştürülmesi işlemidir. Başka bir deyişle görüntü sınıflandırma, uzamsal alandaki mantıksal karar kuralları ya da çok bantlı alanlardaki istatistiksel karar kuralları kullanılarak görüntüdeki piksellerin arazi örtüsü sınıflarına göre kategorize edilmesidir (Gao, 2009). Uzaktan algılamada görüntü sınıflandırma algoritmaları uygulama çeşitliliğine göre geliştirilmektedir. Bu algoritmalar En Küçük Mesafe, En Çok Benzerlik (EÇB) algoritmaları gibi piksel tabanlı yaklaşımlar yaygın olarak kullanılmaktadır. Bu tür yaklaşımlarda görüntünün spektral özellikleri göz önünde bulundurulurken mekânsal bileşenler göz ardı edilir. Bu nedenle bu yaklaşıma alternatif olarak spektral bilginin yanında mekânsal bilginin de kullanıldığı nesne tabanlı yaklaşımlar geliştirilmiştir. Nesne tabanlı sınıflandırma yaklaşımları nesneyi parçalara ayırır ve ardından her bir nesnenin istatistikî özelliklerine göre sınıflandırma işlemi gerçekleştirir (Yıldırım vd. 2005). Son yıllarda uydu görüntüleri üzerinden en doğru ve güvenilir bilgiye hızlı bir şekilde ulaşmak için sınıflandırmaya yönelik öğrenme farklı algoritmalar geliştirilmektedir. Yaygın olarak kullanılan öğrenme tabanlı sınıflandırıcılar arasında Rastgele Orman (RO) sınıflandırma yöntemi, Torbalama (Bagging), Hızlandırma (Boosting), Karar Ağaçları, Yapay Sinir Ağları, Destek Vektör Makineleri, K En Yakın Komşuluk gibi sınıflandırıcılar yer almaktadır. Bu algoritmalar makine öğrenmesi yöntemleri olarak da adlandırılır. Makine öğrenmesi yöntemleri, yeterli büyüklükteki veri ve parametreleri kullanarak kullanıcıların göremediği kuralları ve kısıtlamaları otomatik olarak çıkarabilirler. Bu yöntemler, girdi verileri kullanarak oluşturulan karar kuralları ile yeni veri için en uygun modeli bulmaya çalışırlar.

Öğrenme tabanlı bir sınıflandırıcı olan RO yöntemi, yüksek doğruluk sağlaması, hızlı bir sınıflandırıcı olması nedeniyle araştırmacıların dikkatini çekmektedir (Watts ve Lawrence, 2008; Waske, vd. 2007; Gislason 2004; Pál 2003; Akar, vd. 2010). RO sınıflandırıcısı, gürültü faktöründen etkilenmediğinden ve belirli sabit bir modeli olmadığından birçok ağaç tabanlı algoritmalarından farklı çalışmaktadır (Watts ve Lawrence, 2008). RO' nun sınıflandırma performansını test etmeye ve diğer sınıflandırıcılarla karşılaştırmaya yönelik çalışmalar yapılmaktadır.

Literatürde, arazi örtüsünün belirlenmesi ile ilgili olarak RO sınıflandırıcısının kullanılabilirliğine yönelik birçok çalışma mevcuttur. Watts ve Lawrence (2008)' göre RO algoritması, nesne tabanlı yaklaşıma uygulandığında tarımsal bölgelerin belirlenmesinde yüksek doğruluk sağlamaktadır. Cutler, vd. (2007), çevrebilimi ile ilgili

(ekolojik) verileri kullanarak, RO ve farklı istatistiksel sınıflandırıcılar yardımıyla istilacı bitki türlerini sınıflandırmıştır. Jay (2009), RO' nun hem karışık hem de homojen bitki topluluklarında yer alan bitki türlerini sınıflandırmada başarılı olduğunu söylemiş ve 100 ağaç için % 88.37 oranında genel sınıflandırma doğruluğu elde etmiştir. Waske, vd. (2007), arazi örtüsünü 8 sınıfa (tahıl, orman, otlak, meyve bahçeleri, kanola...) ayırmıştır. Çalışmada RO, EÇB ve basit karar ağacı yöntemlerinden %72.50 'lik oranla daha yüksek genel sınıflandırma doğruluğu sağlanmıştır. Ntoulos' vd. (2009), Landsat 5 TM ve Hyperion görüntülerini kullanarak EÇB yöntemine göre, 5 tarımsal ürün (mısır, pamuk, pirinç, domates ve tütün) sınıflandırmış ve Landsat görüntüsünde %81, hyperion görüntüsünde %91 genel sınıflandırma doğruluğu elde etmiştir. Waske ve Braun (2009), tarımsal alanlar için RO, Hızlandırma gibi öğrenme tabanlı yöntemlerini farklı yıl ve bölgelerdeki SAR görüntüsüne uygulamıştır. Bu çalışma, öğrenme tabanlı sınıflandırıcıların, EÇB sınıflandırıcısından %10 daha iyi sonuç verdiğini göstermiştir. Waske ve Braun, RO sınıflandırıcısının arazi örtüsünün sınıflandırılması için uygun olduğunu ve EÇB ya da basit karar ağaçları yöntemlerinden daha yüksek doğruluk sağladığını söylemektedirler. Ham vd. (2005), RO sınıflandırıcısı ile hyperspektral görüntüler kullanarak arazi örtüsünü sınıflandırmış ve yöntemin, yapılan testler sonucunda iyi sonuç verdiğini belirtmiştir. Prasad vd. (2006), RO, regresyon ağaç analizleri (regression tree analysis), torbalama ağaçları ve çoklu adaptif regresyon eğrileri (multivariate adaptive regression splines) tekniklerini kullanarak bitki örtüsü haritaları oluşturmuştur. Prasad, RO' nun üstün öngörü yeteneğinin mevcut ve gelecekteki uygun habitatları belirlemede başarılı bir şekilde kullanılabileceği görüşünü ortaya koymuştur.

Bu çalışmanın amacı, RO sınıflandırma yönteminin tarım alanlarında yetiştirilen ürün çeşitliliğini tespit etmedeki başarısını test etmektir. RO sınıflandırması sonucunda elde edilen sonuçlar aynı alana uygulanmış ve EÇB sınıflandırma sonuçları ile karşılaştırılarak değerlendirilmiştir. Değerlendirmeler, hata matrisleri yardımıyla genel doğruluklar ve ürün sınıflarına ait hata oranları hesaplanarak parsel tabanlı olarak yapılmıştır. Elde edilen sonuçlar RO sınıflandırmasına ait genel hata oranının (%85.89) EÇB sınıflandırma doğruluğunu (%77.96) yaklaşık %8 oranında artırdığını göstermiştir.

1.1 Rastgele Orman (RO) Sınıflandırıcı

Günümüzde RO algoritması, sınıflandırmada çok iyi performans sergilediği için toplu öğrenme yöntemlerine göre sıklıkla tercih edilmektedir. Son yıllarda geliştirilen RO sınıflandırıcısı, hem hızlı hem de yüksek doğruluk sağlaması yönünden toplu öğrenmede çok iyi iki yöntem olarak bilinen Hızlandırma (Freund and Schapire, 1996) ve Torbalama (Breiman, 1996) yöntemlerine göre avantaj sağlamaktadır.

Öğrenme yöntemleriyle karşılaştırıldığında RO sınıflandırıcısı, özellikle Hızlandırma yöntemine göre, eğitim aşamasında çok daha hızlıdır. Yeterliliği ve doğruluğu ile çok kullanışlı bir sınıflandırıcıdır (Gislason vd., 2006). Hızlandırma yönteminin hesap yükü fazla olduğu için Torbalama yönteminden daha yavaştır. Fakat çoğu durumda bu yöntemden daha doğru sonuçlar verir (Breiman, 2001). Hızlandırma yönteminin, çok yavaş ve gürültüye karşı duyarlı olması, tekrarlı eğitimin olabildiği gibi dezavantajlarına karşın RO, hesapsal olarak Hızlandırma sınıflandırıcısından çok daha basittir, gürültüye karşı duyarlı değildir.

RO, ağaç tipi sınıflandırıcılar topluluğudur. Torbalama yönteminin gelişmiş bir şekli olarak kabul edilebilir. (Breiman,

2001). Breiman ve Cutler (2005) RO'yu, şuan ki algoritmalar arasında doğruluğu eşsiz olan bir sınıflandırıcı olarak tanımlamıştır. Ayrıca hızlı ve belirli bir kalıbı olmayan bir yöntem olduğunu söylemektedir. Bu yöntemde ne kadar istenirse o kadar ağaçla çalışılır (Breiman ve Cutler 2005).

RO yönteminde, $\{h(x, \theta_k) \mid k = 1, \dots\}$ şeklinde ağaç tipi sınıflandırıcılar kullanılmaktadır. Burada, x , girdi verisini; θ_k , rastgele vektörü temsil etmektedir (Breiman, 2001; Breiman, <http://oz.berkeley.edu/users/breiman/randomforests.html>).

Breiman'ın (2001) RO yönteminde, Torbalama, rastgele özellik seçimi ile birlikte ele alınır. RO'da Torbalama yönteminin tercih edilmesinin iki önemli nedeni vardır; birincisi, Torbalama işleminde rastgele özellik kullanıldığından doğruluğunun artması; ikincisi, genelleştirilmiş hataların (Out-of-bag (OOB)) hesaplanmasıdır (Breiman, 2001). Rastgele özellik seçimi için öncelikle gerçek veri setinden yer değiştirmeli olarak yeni bir eğitim veri seti oluşturulur. Ardından, rastgele özellik seçimi kullanılarak yeni eğitim setinden bir ağaç geliştirilir. Geliştirilen ağaçlar budanmaz (Archer, 2008; Breiman, 2001). Pal (2005), budama metodunun seçiminin ve özellik seçim ölçütlerinin olmamasının ağaç tabanlı sınıflandırıcıların performansını etkilediğini belirtmektedir. Budamanın olmaması RO'yu diğer karar ağacı yöntemlerinden daha avantajlı hale getirmektedir.

RO sınıflandırıcısı ile bir ağaç üretmek için kullanıcı tarafından tanımlanan 2 parametre gereklidir. Bu parametreler, en iyi bölünmeyi belirlemek için her bir düğümde kullanılan değişkenlerin sayısı (m) ve geliştirilecek ağaçların sayısı (N)'dir (Pal, 2005).

Kullanıcı tarafından başlangıç m değeri rastgele seçilir sonraki m 'ler genelleştirilmiş hatalara (OBB) göre artırılır ya da azaltılır. m azalınca korelasyon ve güç azalır, m artınca korelasyon ve güç artar. Bu şekilde en uygun m bulunur ve sınıflandırma duyarlılığı artar, hata azalır. Breiman, (2002)'a göre, m değişken değeri seçilirken, m değerinin M (toplam değişken sayısı) değişkeninin kareköküne eşit olarak alınması genellikle optimum sonuçlar verir. RO'da tüm girdi değişkenleri kullanılmamaktadır. RO algoritmasının hesap karmaşıklığını sadeleştirmek ve ağaçlar arasındaki korelasyonu azaltmak amacıyla bir m değeri seçimi yapılmaktadır. Bu m değeri, sınıflandırmada önemli olan değişken sayısını ifade etmektedir. Önemli değişkenler, değişken önem ölçümleriyle belirlenmektedir. RO'dan elde edilen 3 parametre vardır. Bu parametreler genelleştirilmiş hata, değişken önemi (variable importance) ve yakınlık analizidir (proximity analysis) (Chen, 2008).

Genelleştirilmiş hata verisi, sınıflandırma doğruluğunun anlaşılmasına yardımcı olur. T_k eğitim verisinden T_k yer değiştirmeli yeni eğitim verisi üretilir. Yeni eğitim veri seti kullanılarak $h(x, T_k)$ sınıflandırıcısı oluşturulur. Sınıflandırıcı ile çantaya atılmış tahminlerden oylama yapılır. Eğitim verisindeki her x, y için sadece bu sınıflandırıcı ile oylama gerçekleşir. T_k, x ve y 'yi içermez. (Breiman, 2001).

RO algoritması, genelleştirilmiş hata verisindeki verilerin yerleri değiştirildiğinde tahmin hatasının ne kadar olduğunu inceleyerek değişkenlerin önemini, etkilerini (variable importance) hesaplar (Liaw ve Wiener, 2002). Değişken önemi ölçümleri yani kullanılan değişkenlerin ne kadar önemli olduğunun belirlenmesi, değişkenlerin yerleri değiştirilerek yapılır. Değişimler sonucunda oluşan hatalar o değişkenin işlemdeki önemini ortaya koyar. 4 çeşit değişken önemi ölçüm yöntemi vardır. Bunlar; 1: hata artışı (error increase); 2: ortalama hata payı artışı (average margin increase); 3: hata payı

artışının diferansiyeli (differential of margin increases) ve 4: Gini düşüşü (Gini decrease)'dür (Furlanello, 2003). Furlanello (2003)'ün çalışmasında 2. ve 4. yöntem diğerlerinden daha tutarlı sonuçlar vermiştir.

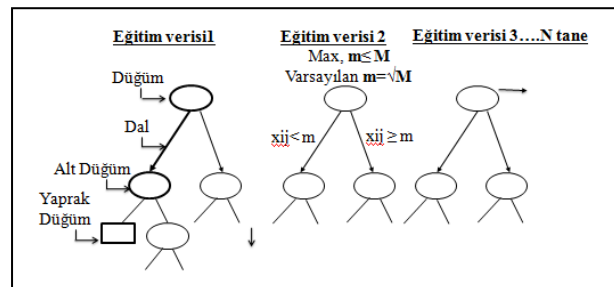
RO, budama olmadan en büyük boyutta ağaç geliştirmek için CART (Classification and Regression Tree) algoritmasını kullanılmaktadır (Breiman, 2001). CART algoritmasında, bir düğümde belirli bir ölçüt uygulanarak bölünme işlemi gerçekleştirilir. Bunun için önce tüm niteliklerin var olduğu değerler göz önüne alınır ve tüm eşleşmelerden sonra iki bölünme elde edilir. Bu bölünmeler üzerinde seçme işlemi uygulanır (Özkan, 2008). Bölünme işlemlerinde homojen sınıf dağılımına sahip düğümler tercih edilir. Düğüm homojenliğinin ölçümünde; Gini İndeks, Entropy, Yanlış Sınıflama Hatası (Misclassification Error), Gain Oranı Kriteri (Gain Ratio Criteria) gibi ölçütler kullanılmaktadır. RO yöntemi, Gini indeksini kullanılmaktadır. Verilen bir T eğitim veri seti için rastgele bir örnek (piksel) seçilsin ve bu örnek C_i sınıfına ait olsun. Bu duruma göre Gini indeksi şöyle ifade edilir (Eşitlik 1);

$$\sum_{j=1}^M (f(C_i, T) / |T|) (f(C_j, T) / |T|)$$

[1]

Eşitliğinde [1], $f(C_i, T) / |T|$ seçilen örneğin C_i sınıfına ait olma olasılığını gösterir (Pal, 2005).

Gini ölçümleriyle, en küçük Gini indeksine sahip olan bölünme pozisyonu belirlenir. (Takçı, 2008). Gini indeksi büyüdükçe sınıf heterojenliği artarken, Gini İndeksi 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). Kaç tane ağaç üretmek istenirse her düğüm için en iyi dal belirlenerek o kadar ağaç üretilir (Liaw and Wiener, 2002). Kısaca oluşturulan eğitim verileri kullanılarak belirlenen bölünme ölçütlerine göre düğümler dallara ayrılmakta ve ağaç yapılırlı oluşmaktadır. Şekil 1'de RO sınıflandırıcısında belirlenen en uygun bölünme pozisyonlarına göre oluşturulan ağaç yapısı örneği gösterilmiştir. Şekildeki x_{ij} , girdi verilerini temsil etmektedir.



Şekil 1. RO yöntemine ait ağaç yapısı

Ağaç gelişiminden sonraki aşama, girdi verisinin sınıflandırılması işlemidir. Bu işlemde, RO, girdi verisini ormandaki her bir ağaç yerleştirir. Belirlenen ağaçlar arasında oylama yapılır ve en çok oyu alan ağaç bir sınıfa atanır (Liaw and Wiener, 2002).

1.2 Çalışma Alanı ve Veri Seti

Çalışma alanı, Bursa ilinin Karacabey ilçesinde yer alan Karacabey Ovasında yer almaktadır (Şekil 2). Yaklaşık 95 km²

lik bir alanı kapsayan tarım alanı, 28°10'18" - 28°18'06" enlemleri ve 40°13'43" - 40°08'37" boylamları arasında bulunmaktadır. Karacabey Ovası zengin toprak yapısı ve ikliminin üretime elverişli olması nedeniyle Türkiye'nin en verimli ovaları arasında yerini almaktadır. Alanda yetiştirilen başlıca ürünler domates, biber, buğday, şeker pancarı, mısır ve pirinç olarak sayılabilir.

Analizler için alanın 23 Temmuz 2004 tarihinde çekilmiş SPOT 5 renkli uydu görüntüsü kullanılmıştır. SPOT 5 uydu görüntüsü 10 m mekânsal çözünürlüğe sahip olup yeşil (0.50-0.59 µm), kırmızı (0.61-0.68 µm), yakın kızıl ötesi (0.79-0.89µm) ve kısa dalga kızıl ötesi (1.58-1.75µm) bantlara sahiptir (Şekil 2). Görüntü, 1120 kolon ve 929 sıradan oluşmaktadır. Görüntü çekim tarihinin, ovadaki ürünlerin gelişmiş olduğu dönem olan yaz aylarında olmasına ve mümkün olduğunca bulutsuz bir günde çekilmesine dikkat edilmiştir. Kullanılan uydu görüntüsü, yer gözlem istasyonundan işlem seviyesi 2A olarak ve "Universal Transverse Mercator" (UTM) koordinat sisteminde alınmıştır.



Şekil 2. Çalışma alanına ait SPOT5 görüntüsü ve vektör veri

SPOT 5 görüntüsüne ek olarak parsel tabanlı sınıflandırma doğruluklarını incelemek amacıyla alanın 1:5000 ölçekli kadastral verilerinden üretilmiş, parsel sınırlarını içeren vektör veri kullanılmıştır (Türker ve Arıkan, 2005). Ürün bilgileri, görüntü çekim tarihi ile eş zamanlı olarak toplanmış ve veritabanına kaydedilmiştir. Analizlerin doğru şekilde yapılabilmesi amacıyla SPOT 5 görüntüsü vektör veri yardımıyla 20 adet yer kontrol noktası kullanılarak geometrik olarak düzeltilmiştir. Geometrik düzeltme işlemi için ikinci derece polinom model ve düzeltme sonrasında piksellere atanacak yeni radyometrik değerler için en yakın mesafe yöntemi kullanılmıştır. Yapılan geometrik düzeltme sonucunda ortaya çıkan hata miktarı $\pm 0,41$ piksel olarak hesaplanmıştır (Özdarıcı, 2005).

2. YÖNTEM

2.1 Rastgele Orman Sınıflandırma Yöntemi

RO algoritması ile görüntünün sınıflandırılması için eğitim verilerinin oluşturulması gerekmektedir. Bu işlem için ERDAS Imagine programı kullanılmıştır. Öncelikli olarak belirlenen 5 sınıfı yansıtabilecek şekilde eğitim (train) ve test verileri toplanmış ve bir eğitim dosyası oluşturulmuştur. Bu maksatla, her sınıftan eşit sayıda örnek alınarak sınıfları iyi temsil edecek toplam 9684 piksel toplanmıştır. Oluşturulan veriden eğitim ve test verilerinin üretilmesinde ve RO sınıflandırma işleminde

Jaiantila (2009)'a ait Matlab kodundan yararlanılmıştır. Çalışma kapsamında bu matlab kodu düzenlenmiş ve rastgele seçim ile birbirinden bağımsız olarak 5000 piksel eğitim verisi için, 4684 piksel ise test verisi için hazırlanmıştır. Eğitim verisi görüntüyü sınıflandırmak için, test verisi ise sınıflandırma modelini test etmek için kullanılmaktadır.

RO sınıflandırıcısı için ağaç sayısı (N) ve m değişkeni olmak üzere 2 faktör sınıflandırma doğruluğunu önemli derecede etkilemektedir. Bu parametreler kullanıcının tercihine sunulmuştur. Bunun için sınıflandırma modeline en uygun ağaç sayısı ve m değerinin tanımlanması gerekmektedir. En uygun parametrelerin bulunması işlemi için tekrarlı denemeler yapılmıştır. Tablo 1'de görüldüğü gibi ağaç sayılarına göre değişkenlerdeki değişimlerin sınıflandırmayı ne derece etkilediği test edilmiştir. Eğitim verisinde, görüntünün özelliğine göre Band1, Band2, Band3 ve Band4 olmak üzere toplam 4 değişken bulunmaktadır. M toplam girdi değişkenlerinden rastgele m değişkeni $m \leq M$ olacak şekilde her bir düğüm için seçilir (Breiman ve Cutler, 2005). Varsayılan değer $m = \sqrt{M}$ 'dir (Gislason, et al. 2004). Bu nedenle m değerleri 1 ile 4 arasında seçilmiştir.

Tablo1. RO sınıflandırıcısının seçilen ağaç sayıları ve m

değişkenlerine göre OOB hataları, test doğrulukları ve işlem süreleri

N	m	OOB Hatası (%)	Test Doğruluğu (%)	Kappa	İşlem Süresi (sn)
100	1	97.48	97.33	0.9719	105.62
	2	97.63	97.31	0.9717	107.57
	3	97.57	97.29	0.9714	108.38
	4	97.69	97.27	0.9712	114.99
200	1	97.54	97.28	0.9714	150.41
	2	97.61	97.61	0.9748	155.66
	3	97.20	97.46	0.9733	165.47
	4	97.35	97.48	0.9734	168.49
250	1	97.14	97.46	0.9733	178.37
	2	97.40	97.72	0.9758	188.60
	3	97.82	97.59	0.9746	188.28
	4	97.50	97.42	0.9728	195.47
500	1	97.08	97.18	0.9708	305.45
	2	97.54	97.35	0.9721	324.34
	3	97.78	97.23	0.9708	348.34
	4	97.42	97.37	0.9723	331.27

Seçilen parametrelere göre hem işlem süreleri hem de test doğrulukları incelendiğinde RO sınıflandırması için en uygun ağaç sayısı 200 ve m parametresi de 2 olarak belirlenmiştir.

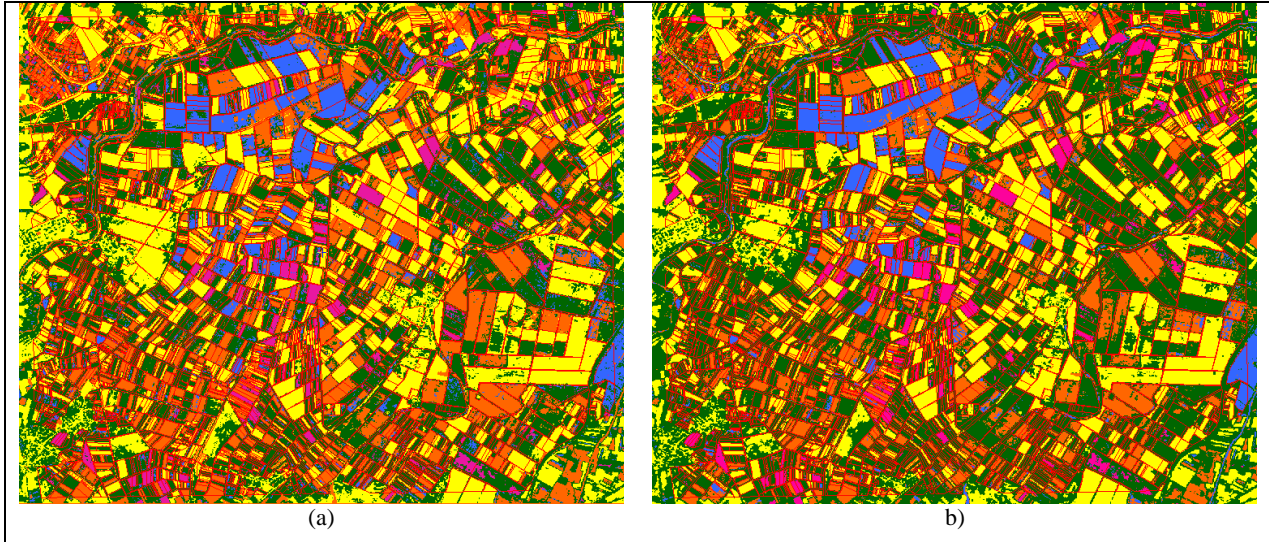
Tablo 2'de görüldüğü gibi test verileri ile yapılan sınıflandırma sonucunda RO yönteminin test doğruluğu 97.61 olarak elde edilmiştir.

Tablo 2. RO Sınıflandırma yönteminde kullanılan test verisine ait hata matrisi

	Sınıflar													Satır Toplamı	Üretici Doğruluğu(%)	Kullanıcı Doğruluğu(%)
	1	2	4	5	7	10	11	12	15	17	24	30	31			
1	551	0	0	0	0	0	0	0	0	0	0	0	0	551	100%	100%
2	0	180	0	0	0	0	0	0	0	0	0	0	0	180	100%	100%
4	0	0	297	0	0	0	0	0	0	0	16	0	0	313	96%	95%
5	0	0	0	331	1	12	0	0	0	0	0	0	0	344	95%	96%
7	0	0	0	3	307	0	0	0	0	0	0	0	0	310	100%	99%
10	0	0	0	9	0	451	0	0	0	0	0	0	5	465	96%	97%
11	0	0	0	0	0	0	520	0	0	0	0	0	0	520	100%	100%
12	0	0	0	0	0	0	0	521	0	0	0	0	0	521	100%	100%
15	0	0	0	0	0	0	0	0	442	0	0	0	0	442	99%	100%
17	0	0	0	0	0	0	0	0	0	246	0	4	9	259	95%	95%
24	0	0	13	0	0	0	0	0	3	0	230	2	0	248	93%	93%
30	0	0	0	1	0	0	0	0	0	3	0	435	3	442	97%	98%
31	0	0	0	3	0	6	0	1	0	11	0	7	61	89	78%	69%
Sütun Toplamı	551	180	310	347	308	469	520	522	445	260	246	448	78	4684		
	Genel Doğruluk= 97,61%															
	Kappa= 0,9748															

Test verileri ile yapılan sınıflandırma doğruluğunun yüksek olması, RO yönteminin görüntünün tamamı üzerinde uygulanmasına olanak sağlamıştır. Aynı görüntü, aynı eğitim alanları kullanılarak EÇB sınıflandırma yöntemine göre de sınıflandırılarak elde edilen sonuçlar karşılaştırılmıştır. Hem RO hem de en çok benzerlik yöntemleri ile elde edilen sınıflandırılmış görüntüler Şekil 3'de gösterilmektedir.

Buna göre 200 ağaç sayısı ve $m=2$ değeri kullanılarak yapılan RO sınıflandırmasının en yüksek genel doğruluk (%85.89) ve kapa (%0.7977) oranlarını sağladığı gözlenmiştir. Farklı parametreler kullanılarak elde edilen sonuçların da en yüksek sonuca yakınlığı dikkat çekmektedir.



Şekil 3. (a) RO ve (b) EÇB yöntemi ile sınıflandırılmış görüntüler ve vektör veri

2.2 Parsel Tabanlı Doğruluk Analizi

Görüntülerin sınıflandırma sonrası genel doğrulukları belirlemek amacıyla parsel tabanlı bir doğruluk analizi tekniği uygulanmıştır. Bunun için ilk olarak görüntü RO ve EÇB sınıflandırma yöntemi yardımıyla piksel tabanlı olarak sınıflandırılmış ve ardından parsel içine düşen piksellere ait en çok tekrarlanan piksel değerleri (mode) hesaplatılarak bu değer parsel etiket olarak atanmıştır. Doğruluk ölçümleri sırasında alandaki ürün bilgilerinin görüntü çekim tarihi ile eş zamanlı olarak veri tabanında tutulduğu 1021 adet parsel kullanılmıştır. Çalışmada parsel tabanlı doğruluk analizi tekniğinin kullanılmasının nedeni, alandaki her bir parselin tek ürün içerdiğinin varsayımından yola çıkılmasıdır. Tablo 3'de RO yöntemine ait çeşitli parametreler kullanılarak elde edilmiş parsel tabanlı analiz sonuçları verilmektedir.

Tablo 3. RO için Parsel Tabanlı Doğruluk Analizi Sonuçları

<i>N</i>	<i>m</i>	Genel Doğruluk (%)	Kappa
100	1	83.84	0.772215
100	2	85.40	0.791343
100	3	85.60	0.793430
100	4	85.02	0.784789
200	1	85.41	0.791630
200	2	85.89	0.797777
200	3	85.80	0.796318
200	4	84.70	0.781095

RO ve EÇB yöntemlerine ait hata matrisi sonuçları Tablo 4'de verilmektedir.

Tablo 4. RO (a) ve EÇB (b) yöntemleri sonuçları

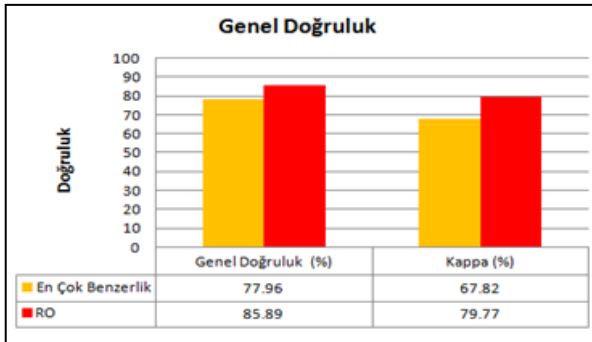
	Cr	Tm	Rc	Rs	Sb	RTot
Cr	170	8	0	7	0	185
Tm	74	341	1	17	5	438
Rc	12	1	26	0	0	39
Rs	6	6	0	308	0	320
Sb	0	5	2	0	32	39
CTot	262	361	29	332	37	1021
Producer's(%)	0.64	0.94	0.89	0.92	0.86	
User's (%)	0.91	0.77	0.66	0.96	0.82	
Overall (%): 85.89	Kappa (%): 79.77					

(a)

	Cr	Tm	Rc	Rs	Sb	RTot
Cr	108	4	0	2	0	114
Tm	143	351	4	44	9	551
Rc	5	0	23	0	0	28
Rs	6	2	0	286	0	294
Sb	0	4	2	0	28	34
CTot	262	361	29	332	37	1021
Producer's(%)	0.41	0.97	0.79	0.86	0.75	
User's (%)	0.94	0.63	0.82	0.97	0.82	
Overall (%): 77.96	Kappa(%): 67.82					

(b)

RO ve EÇB yöntemlerine ait Parsel Tabanlı Doğruluk Analizi Sonuçları şekil 4.'te grafiksel olarak gösterilmiştir.



Şekil 4. RO ve EÇB yöntemlerine ait Parsel Tabanlı Doğruluk Analizi Sonuçları

3. SONUÇLAR

Çalışmada Karacabey Ovası'nda yetiştirilen 5 farklı ürün sınıfı RO sınıflandırma yöntemi yardımıyla SPOT 5 görüntüsü kullanılarak sınıflandırılmıştır. En güvenilir sonuca ulaşabilmek için görüntü üzerinde çeşitli parametreler test edilmiştir. Ardından SPOT 5 uydu görüntüsü aynı örnek alanlar kullanılarak EÇB yöntemi ile sınıflandırılmış ve bu sınıflandırma sonucunda elde edilen tematik haritanın doğruluğu RO yöntemi sonucunda üretilen harita ile karşılaştırılmıştır. Elde edilen tematik haritaların değerlendirilmesinde hata matrislerinden faydalanılmıştır. Bu

matrisler sayesinde ürünlere ait doğruluklar, genel doğruluk ve kappa oranları hesaplanarak ürün sınıfları için en güvenilir yöntem belirlenmeye çalışılmıştır. Sonuçlar, RO sınıflandırma yöntemine ait genel doğruluk ve kappa oranlarının sırasıyla 85.89% ve 77.96% olduğunu göstermiştir. Diğer taraftan EÇB yöntemi sonucunda elde edilen genel doğruluk oranı 79.77% olarak hesaplanırken kappa oranı bu sınıflandırma yöntemi için 67.82% olarak bulunmuştur. Buna göre RO sınıflandırma yöntemi EÇB sınıflandırma yöntemi sonucunu yaklaşık %8 oranında arttırmıştır. Farklı parametreler kullanılarak yapılan RO sınıflandırma sonuçlarının da birbirleri ile tutarlı olması RO sınıflandırma yönteminin tarım alanlarındaki ürün çeşitliliğini tespit etmedeki başarısını dikkate değer kılmakta ve bu yöntemin farklı test alanlarındaki etkilerini değerlendirmenin önemini ortaya koymaktadır.

Kaynaklar:

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, pp.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 (15.11.2009).

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 (06.11.2009)

Breiman, L., 2001, Random Forests, Machine learning, 2001 Kluwer Academic Publishers, 45(1), 5-32.

Breiman, L., 1996. Bagging predictors. Machine Learning 26(2), 123-140.

Breiman, L. Random Forests-Random Features, Technical report, <http://oz.berkeley.edu/users/breiman/randomforests.html>, (06.11.2009).

Chen, C. H., 2008, *Image Processing For Remote Sensing*, CRS Press, Taylor&Francis Group, USA.

Christian B., Krishnappa N. S. R., 2009, Classification of tropical trees growing in a sanctuary using Hyperion (EO-1) and SAM algorithm, *Current Science*, 96(12), 1601-1607.

Daughtry C.S.T., Doraiswamy P.C., Hunt E.R., Jr., Stern A.J., McMurtrey J.E., III, Preuger J.H.. 2006. Remote Sensing Of Crop Residue Cover And Soil Tillage Intensity, Soil and Tillage Research 91(1-2):101-108.

De Jong, S. M, Van Der Meer, F. D, 2005, *Remote Sensing Image Analysis: Including the Spatial Domain*, Springer Science, Volume 5, United States of America.

Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. In: Machine Learning. *Proceedings of the Thirteenth International Conference*. pp. 148-156.

Furlanello, C., M. Neteler, S. Merler, S. Menegon, S. Fontanari, A. Donini, A. Rizzoli, and C. Chemini, 2003. GIS And The Random Forests Predictor: Integration In R For Tick-Borne Disease Risk Assessment, *Proceedings of the 3rd International*

- Workshop on Distributed Statistical Computing (DSC 2003), Vienna, Austria, URL: <http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/FurlanelloEtAl.pdf> (20.11.2010).
- 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(2006), 294-300.
- Gislason, P.O., Benediktsson, J.A, Sveinsson, J.R., 2004, Random forest classification of multi-source remote sensing and geographic data, *Geoscience and Remote Sensing Symposium, 2004, IGARSS '04, Proceedings, 2004 IEEE International*, vol.2, 1049 – 1052.
- Ham, J., Y. Chen, M.M. Crawford, and J. Ghosh, 2005. Investigation Of The Random Forest Framework For Classification Of Hyperspectral Data, *IEEE Trans. Geosci. Remote Sens.* v43. 492-501.
- Huang, Y., Fipps, G., 2006, Landsat satellite multi-spectral image classification of land cover change for gis-based urbanization analysis in irrigation districts: evaluation in Low RioGrandeValley1, <http://idea.tamu.edu/documents/LandsatReport06Aug.pdf> (21.02.2010).
- IDRISI Klimanjaro, 2003, Guide to GIS and Image Processing Volume 2, Idrisi Production, Clark Labs, USA, pp.57-82
- Jaiantila, A., 2009, Classification Random Forest, [http://code.google.com/p/randomforestmatlab/downloads/detail?name=RF_MexStandalone-v0.02.zip&can=2&q=\(25.08.2009\)](http://code.google.com/p/randomforestmatlab/downloads/detail?name=RF_MexStandalone-v0.02.zip&can=2&q=(25.08.2009)).
- Jaiantila, A., 2009, Classification Random Forest, [http://code.google.com/p/randomforestmatlab/downloads/detail?name=RF_MexStandalone-v0.02.zip&can=2&q=\(25.08.2009\)](http://code.google.com/p/randomforestmatlab/downloads/detail?name=RF_MexStandalone-v0.02.zip&can=2&q=(25.08.2009)).
- Jay, S., Lawrence, R., Repasky, K., Keith, C., 2009, Invasive Species Mapping Using Low Cost Hyperspectral Imagery, *ASPRS 2009 Annual Conference Baltimore*, Maryland ♦ March 9-13, 2009.
- Liaw A., Wiener M., 2002, Classification And Regression By Random Forest, *R News*, Vol.2/3, December.
- Ntouros, K. D., Gitas, I. Z., Silleos, G. N., 2009, Mapping Agricultural Crops With Eo-L Hyperion Data, 978-1-4244-4687-2/09/\$25.00 ©2009 IEEE.
- Özdarıcı, A., 2005, “Comparison Of Different Resolution Images For Polygon-Based Crop Mapping”, Yüksek Lisans Tezi, ODTÜ Fen Bilimleri Enstitüsü, Jeodezi ve Coğrafi Bilgi Teknolojileri, Ankara.
- Özkan, Y., 2008, *Veri Madenciliği Yöntemleri*, Papatya Yayıncılık, İstanbul.
- Pal, M., 2005, Random Forest Classifier For Remote Sensing Classification, *International Journal Of Remote Sensing*, 26(1) , 217-222.
- Pal, M., 2003, Random Forest For Land Cover Classification, *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, vol.6, 3510 – 3512.
- Prasad, A.M., L.R. Iverson, and A. Liaw, 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction, *Ecosystems*, 9:181–199.
- Sajjad A., Ajay K., Haroon S., 2010, Estimating soil moisture using remote sensing data: A machine learning approach, *Advances in Water Resources* 33 (2010) 69–80
- Takçı, H., 2008, *Veri Madenciliği Dersi Notları*, (Veri Madenciliği Dersi – GYTE – Dr. Hidayet Takçı 10/05/2008)
- Tso, B., Mather P. M., 2009. *Classification Methods For Remotely Sensed Data*, Second Editon, Taylor & Francis Group, United States of America.
- Türker M. ve Arıkan M., 2005, Sequential masking classification of multi-temporal Landsat7 ETM+ images for field-based crop mapping in Karacabey, Turkey, *International Journal of Remote Sensing* Vol. 26, p. 3813–3830.
- Villalon-Turrubiates, I.E.. 2009, Remote Sensing Signatures Extraction For Hydrological Resources Management Applications, AICCSA, 2009 *IEEE/ACS International Conference on Computer Systems and Applications*, pp.567-570.
- Waske, B., Braun, M., 2009, Classifier Ensembles For Land Cover Mapping Using Multitemporal SAR Imagery, *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (2009) 450_457
- Waske, B., Heinzl, V., Braun M., Menz, G., 2007, Random Forests For Classifying Multi-Temporal Sar Data, Proc. ‘*Envisat Symposium 2007*’, Montreux, Switzerland, 23–27 April 2007 (ESA SP-636, July 2007) <http://envisat.esa.int/envisatsymposium/proceedings/sessions/3D3/461589wa.pdf> (13.03.2010).
- 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 (2011) 66–75
- Watts, J. D. and R. L. 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*. Vol. XXXVII. Part B7. Beijing 2008.
- Yıldırım, İ., Ersoy, O.K., Yazgan B., 2005, Improvement of classification accuracy in remote sensing using morphological filter, *Advances in Space Research* 36 (2005) 1003–1006.
- Zhang W., Xue X., Sun Z., Guo Y., Chi M., Lu H., 2007, Efficient feature extraction for image classification, *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp.1-8