

EŞ DİZİMLİLİK MATRİSİ VE RASTGELE ORMAN SINIFLANDIRICISI İLE ÇAY VE FINDIK ALANLARININ SINIFLANDIRILMASI

Ö. Akar^{a,*}, O. Güngör^a

^a KTU, Harita Mühendisliği, 61080 Trabzon, Türkiye - (ozlemerden, oguzgungor@gmail.com)

ANAHTAR KELİMELER: WorldView-2, Sınıflandırma, Eş Düzlemlilik Matrisi, Rastgele Orman, Doğruluk

ÖZET:

Çalışmanın amacı Doğu Karadeniz Bölgesinde yaygın olarak yetiştirilen ve bölge için ekonomik değere sahip birbirine yakın spektral özellikteki çay ve fındık alanlarını Eş Dizimlilik (Co-occurrence) Matris ve Rastgele Orman (RO) sınıflandırıcısı ile belirlemektir. Yakın spektral özellikteki ürünleri sadece spektral özelliklerine bakarak ayırmak oldukça güçtür. Bu problemi çözmek için çalışmada, spektral özelliğin yanında doku özelliğinin de kullanımının sınıflandırma başarısına katkısı araştırılmıştır. Çalışma alanı olarak hem fındık hem de çay ürünlerinin yetiştirildiği Trabzon ili Sürmene ilçesi Kumru Mahallesi seçilmiştir. 2012 yılına ait 8 bantlı MS (multispektral) ve PAN (pankromatik) WorldView-2 uydu görüntüleri çalışmada kullanılmıştır. Görüntüler Sayısal Yükseklik Modeli, RPC ve yer kontrol noktaları ile geometrik olarak düzeltilmiştir. Daha sonra sınıflandırma işleminde kullanılacak sınıflar için çalışma bölgesinde örnek alanlar toplanmıştır. Bu örnek alanlara göre MS uydu görüntüsünden 7 farklı sınıf (orman, fındık, çay, toprak, gölge, kentsel alan1 ve kentsel alan2) için eğitim pikselleri toplanmış ve sonra eğitim ve test verileri oluşturulmuştur. Bu veriler RO sınıflandırma yöntemine girdi olarak kullanılmıştır. RO öncelikli olarak ürünlerin sadece MS uydu görüntüsünden alınan spektral değerleri dikkate alınarak RO sınıflandırıcı ile sınıflandırmaları yapılmış ve genel sınıflandırma doğruluğu %79.05 olarak elde edilmiştir. Spektral özelliklerin yanında Eş Dizimlilik Matrisi ile elde edilen doku özellikleri kullanıldığında genel sınıflandırma doğruluğu %84.08 olarak elde edilmiştir. Sonuçlar doku özelliğinin RO' nun sınıflandırma doğruluğunu %5 oranında artırdığını göstermektedir. Ayrıca ürünlerden elde edilen doğruluklar da irdelendiğinde fındık alanları için ortalama %13, çay alanları için ise ortalama %7 oranında artış görülmüştür. Elde edilen sonuçlar doku özelliğinin sınıflandırmadaki başarısını ortaya koymaktadır.

CLASSIFICATION PLANTATIONS OF TEA AND HAZELNUT WITH CO-OCCURRENCE MATRIX AND RANDOM FOREST CLASSIFIER

ABSTRACT:

The aim of the study is to determine plantations of hazelnut and tea that have spectrally similar properties that have very valuable source of income as which are commonly grown in the East Blacksea Region with Co-Occurrence Matrix and Random Forest classifier. The crops with similar spectral properties is difficult to distinguish from each other according to only spectral properties. To solve this problem contribution of usage texture features as well as spectral properties to classification has investigated in this study. As study area, Kumru district in Trabzon-Sürmene that grown both hazelnut and tea crops was selected. In the study MS (multispectral), which have eight bands, and PAN(panchromatic) WorldView-2 satellite images were used. The images were geometrically rectified with Digital Elevation Model, RPC and ground control points. Then sample areas was collected for classes used in the classification process in the study area. According to these sample areas, training pixels were collected on MS image for seven class (forest, hazelnut, tea, soil, shadow, urban structure 1 and urban structure 2) and then train and test data were created. These data were utilized as an input to the RF classification. As a priority, taking into account only spectral values of the crops taken from MS satellite images, these crops were classified with RO classifier and overall classification accuracy was obtained as 79.05%. When used texture features generated from co-occurrence matrix as well as spectral properties, overall classification accuracy was obtained as 84.08%. The results showed that texture features improved the accuracy of classification of RO around 5%. Furthermore when user and producer accuracies of crops were investigated, it was shown increased around 13% for plantation of hazelnut, around 7% for plantation of tea. Obtained results reveal the success of texture feature in classification.

KEY WORDS: WorldView-2, Classification, Co-occurrence Matrix, Random Forest, Accuracy

* Corresponding author

1. GİRİŞ

Dünyada gelir kaynakları arasında önemli bir yeri olan tarımsal ürün ve alanların tespitinde de uzaktan algılama tekniklerine sıklıkla başvurulmaktadır. Özellikle rekolte tahmini, tarımsal ürünlerin tespiti, tarımsal alanlardaki erozyonun tespiti, zarar görmüş ürünlerin tespiti ve tarımsal bölgelerin yeryüzünde kapladığı alanların belirlenmesi gibi çalışmalarla bu alanların izlenmesi ve kontrolü yapılarak üretim sırasında meydana gelen her türlü zarar ortadan kaldırılabilecektir. Böylece tarımsal ürünlerin üretiminde yüksek verim elde edilebilecektir. Bu ürün ve alanların daha doğru ve hızlı olarak tespiti için farklı konumsal ve spektral çözünürlükteki görüntülerin yanı sıra görüntüde araziyi temsil eden piksellerin gri değerlerini ve görüntü üzerindeki objelerin şekil, doku özelliklerini dikkate alan obje çıkarımı, filtreleme, sınıflandırma gibi farklı uzaktan algılama teknikleri kullanılmaktadır. Sınıflandırma işlemi en yaygın kullanılan yöntemlerden biridir.

Görüntü sınıflandırma mevcut spektral verileri kullanarak belirli bir spektral sınıfın parçası olan pikselleri etiketleme işlemidir (Richards ve Jia, 2006). Sınıflandırma işlemi iki adımdan oluşur. Birinci adım, objelerin kategorilerini tanımlamaktır. Arazi yüzeyi için düşünüldüğünde bu kategoriler, su yapıları, ağaçlık alanlar, çayır-çimen alanları ... gibi farklı arazi örtüsü türlerini içerir. Sınıflandırma işleminde ikinci aşama ise sınıflandırılan girdilerin etiketlenmesi işlemidir. Sınıflandırılmış tematik haritalardaki bu etiketler nümeriktir. Örneğin su sınıfına ait olan bir piksel, sınıf 1 olarak ya da ağaçlık alana ait olan bir piksel ise sınıf 2 olarak etiketlenebilir. Bu işlemlere sınıflandırma ve tanımlama (etiketleme) denmektedir (Mather, 2004).

Uzaktan algılamada kontrollü görüntü sınıflandırma algoritmaları parametrik ve parametrik olmayan algoritmalar olarak kategorize edilebilir. Parametrik olan algoritmalar eğitim işleminde, eğitim verisinden elde edilen istatistiksel parametreleri (varyans, kovaryans, ortalama,... gibi) kullanır. En Küçük Mesafe, En Çok Benzerlik (EÇB) ve Fisher Lineer Diskriminant sınıflandırıcıları parametrik sınıflandırıcılar arasında yer alır. Parametrik olmayan algoritmalarından 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 metotları olarak da bilinir. İstatistiksel (parametrik) 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).

Literatürde, arazi örtüsünün belirlenmesi, tarımsal ürünlerin tespiti gibi tarımsal çalışmalarda RO sınıflandırıcısının kullanılabilirliğine ve doğruluğuna yönelik birçok çalışma mevcuttur. Watts ve Lawrence (2008)' e göre RO algoritması, obje tabanlı yaklaşıma uygulandığında tarımsal bölgelerin belirlenmesinde yüksek doğruluk sağlamaktadır. Cutler, vd. (2007), ekolojik verileri kullanarak, RO ve başka 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. Ok vd. (2012), SPOT 5 uydu görüntüsü üzerinden 5 çeşit tarımsal ürünü (mısır, domates, biber, pirinç, şeker pancarı ve buğday) RO ve En Çok Benzerlik sınıflandırıcıları ile sınıflandırmıştır. Çalışmasında ürün sınıfları için RO sınıflandırma yöntemi sonucunda elde edilen genel ortalama hata oranını % 85.89,

EÇB sınıflandırma yöntemi kullanılarak üretilen genel ortalama hata oranının % 77.96 olduğunu belirtmiştir. Buna göre RO sınıflandırma yönteminin EÇB sınıflandırma yöntemi sonucunu yaklaşık % 8 oranında artırdığı gözlenmiştir. Bu oran, RO sınıflandırma yönteminin tarım alanlarında yetiştirilen ürün çeşitliliğini tespit etmede önemli bir yöntem olabileceğini göstermektedir. Bunun yanında literatürde doku analizi yardımıyla düzenli dikim alanlarının bulunmasına yönelik çalışmalar da yapılmaktadır. İlgili çalışmalarda, Eş dizimlilik (co-occurrence) ve Fourier analizi tabanlı yöntemler sıklıkla kullanılmaktadır (Yalnız ve Aksoy, 2009). Aksoy (2009), yüksek çözünürlüklü pankromatik görüntü üzerinde düzenli bir ekim alanı olarak görülen fındık alanlarının doku özelliğinden yararlanılmış ve yöntem olarak Laplacian of Gaussian nokta filtresini kullanmıştır. Çalışma, IKONOS ve QuickBird görüntüleri için seçilen test alanları ile meyve bahçelerinin iyi bir şekilde belirlendiğini göstermektedir.

Bu çalışma, özellikle ekonomik değeri yüksek ve spektral olarak birbirine yakın özelliğe sahip olan çay ve fındık alanlarının uydu görüntüleri ve doku özellikleri kullanılarak RO sınıflandırıcısının performansını artırmak ve bu ürünlerin sağlıklı bir şekilde birbirinden ayırt edilmesini amaçlamaktadır. Bildiride ilk olarak; sınıflandırma işlemi için kullanılan RO algoritması ve doku çıkarımı için kullanılan Eş dizimlilik yöntemi hakkında teorik bilgi verilmiştir. Sonraki bölümlerde çalışma alanına ve kullanılan verilere değinilmiş ve yapılan uygulama anlatılmıştır. Son olarak da elde edilen sonuçlara yer verilmiştir.

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 ve 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 olabilmesi 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, 2009).

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 doğruluğun artması; ikincisi, genelleştirilmiş hataların (Out-of-bag (OOB)) hesaplanmasıdır (Beriman, 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; Beriman, 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 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. (Beriman, 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 vd., 2003). Furlanello (2003)'ün çalışmasında 2. ve 4. yöntem değerlerinden daha tutarlı sonuçlar vermiştir.

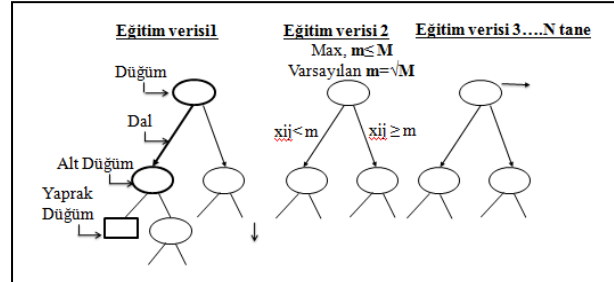
RO, budama olmadan en büyük boyutta ağaç geliştirmek için CART (Classification and Regression Tree) algoritmasını kullanmaktadır (Beriman, 2001). CART algoritmasında, bir düğümde belirli bir ölçüt uygulanarak bölünme işlemi gerçekleştirilir. Bunun için önce tüm niteliklerin var olduğu değerler göz önüne 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 kullanmaktadır. 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);

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

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 indeksinden daha az olduğunda o dal başarılıdır. Gini indeksi sıfıra ulaştıkça 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 ve Wiener, 2002). Kısaca oluşturulan eğitim verileri kullanılarak belirlenen bölünme ölçütlerine göre düğümler dallara ayrılmakta ve ağaç yapıları 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

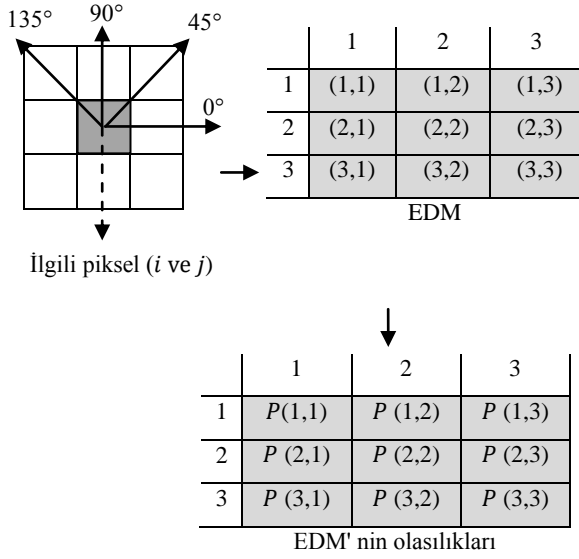
Ağaç gelişiminden sonraki aşama, girdi verisinin sınıflandırılması işlemidir. Bu işlemde, RO, girdi verisini ormandaki her bir ağaca yerleştirir. Belirlenen ağaçlar arasında oylama yapılır ve en çok oyu alan ağaç bir sınıfa atanır (Liaw ve Wiener, 2002).

1.2 Eşdizimlilik Matrisi

Eş Dizimlilik Matrisi (EDM) görüntüdeki gri değerlerinin birbirleriyle olan ilişkisel bağlantılarını kullanarak doku bilgisini elde etmeyi amaçlar (Tso ve Mather, 2009). Genel anlamda görüntüde tekrar eden piksel çiftlerinin sayısı hakkında bilgi verir.

EDM yöntemi ile doku bilgisinin çıkarımı işlemi şu şekilde gerçekleşmektedir. Öncelikle 3x3, 5x5 gibi bir pencere tanımlanır. Bu pencere içerisindeki piksellerin gri değerleri seçilen gri değer sayısına göre ölçeklenir. EDM' nin oluşturulması için gri seviye sayısı, yön ve mesafe gibi parametreler tanımlanmalıdır. Gri seviye sayısı, EDM' nin boyutunu belirler. Örneğin gri değerlerin 0 ile 255 arasında değiştiği bir görüntüde gri değer sayısı 2^n olmak üzere, EDM' nin boyutu $n \times n$ yani 8x8 olarak seçilirse ve bu gri değerler 1 ile 8 arasına ölçeklenir. Burada amaç, gri değer sayısını azaltan matris oluşturmak. Oluşturulan matris her zaman karedir ve boyutu 8x8, 16x16... gibi seçilebilir. Yön parametresi ise,

piksel çiftlerinin yönünü tanımlamada kullanılır. Tanımlanan 0°, 45°, 90° ve 135° gibi yönlerdeki piksel çiftleri matrise işlenir.



Şekil 2. EDM'nin oluşturulması

Şekil 2'de de görüldüğü gibi i ve j komşu iki pikselin gri tonları olmak üzere her (i, j) için bu işle yapılır ve matris bu şekilde oluşturulur. Sonraki aşamada her bir (i, j) matris içerisindeki toplam sayıya bölünerek olasılıklar hesaplanır. Her bir $P(i, j)$ yardımıyla da görüntü üzerinde tanımlanan pencerenin ilgili pikselinin ortalama, varyans, entropi, homojenlik, kontrast ve korelasyon... gibi istatistiksel bilgileri hesaplanır.

$$\text{Benzerlik; } \sum_i \sum_j P_{ij}^2 \quad [2]$$

$$\text{Entropi; } - \sum_i \sum_j P_{ij} \log(P_{ij}) \quad [3]$$

$$\text{Kontrast; } \sum_i \sum_j (i - j) P_{ij} \quad [4]$$

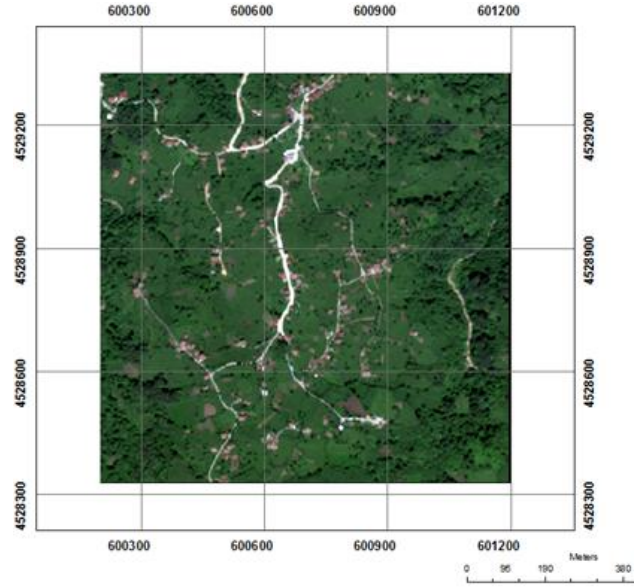
$$\text{Homojenlik; } \sum_i \sum_j \frac{P_{ij}}{1 + |i - j|} \quad [5]$$

$$\text{Korelasyon; } \sum_i \sum_j \frac{(i - \mu)(j - \mu) P_{ij}}{\sigma^2} \quad [6]$$

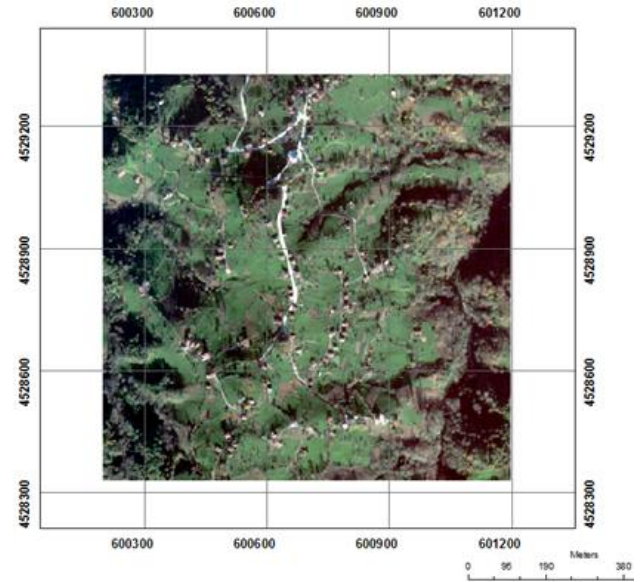
P_{ij} , ij değerlerinin olasılığını, μ ortalama değerini ve σ standart sapma değerini ifade etmektedir.

1.3 Çalışma Alanı ve Veri Seti

Çalışma alanı olarak hem fındık hem de çay ürünlerinin yetiştirildiği Trabzon ili Sürmene ilçesi Kumru Mahallesi seçilmiştir. Sürmene dünya üzerinde kuzey yarım kürenin doğu yarısında, dünyanın kuzeydoğusunda bulunan Anadolu yarımadasının Karadeniz bölgesinin Doğu Karadeniz bölümünde 40° 07' doğu boyları ve 40° 52' kuzey enlemi üzerinde yer alır. İlçe nüfusunun % 80'i tarımla uğraşmaktadır.



a)



b)

Şekil 3 Çalışma Alanları a) Yaz dönemine ait Worldview-II MS görüntüsü (5,3,2), b) Kış dönemine ait Worldview-II MS görüntüsü (5,3,2)

En çok gelir getiren ürün çay ve fındıktır. Çalışma alanı farklı dönemlerde gelişim gösteren çay ve fındık gibi tarımsal ürünleri bünyesinde barındırdığı için bu ürünleri birbirinden daha iyi ayırt etmek adına yaz ve kış dönemlerinde çekilen görüntüler 111Y296 nolu TÜBİTAK projesi kapsamında temin edilmiş ve bu çalışma kapsamında kullanılmıştır. Kullanılan uydu görüntüleri, 2012 yılına ait 2m konumsal çözünürlüklü 8 bantlı MS (multi spektral) ve 0.5m konumsal çözünürlüklü PAN (pankromatik) WorldView-II uydu görüntüleridir. Pankromatik bandı 450-800nm dalga boyuna sahipken, Sahil (coastal) bandı 400-450nm, mavi bant 450-510nm, yeşil bant 510-580nm, sarı bant 585-625nm, kırmızı bant 630-690nm, Kırmızı Kenar bandı 705-745nm, Yakın IR1(NIR1) bandı 770-895nm ve Yakın IR2 (NIR2) bandı 860-900nm dalga boyuna sahiptir.

Görüntüler öncelikli olarak geometrik düzeltme işlemi için RPC verileri, 1/25000lik haritadan üretilen Sayısal Yükseklik Modeli (DEM), yer kontrol noktaları kullanılmış ve geometrik düzeltme modeli olarak da WorldView RPC modeli seçilmiştir.

MS görüntüsünün geometrik düzeltme işleminde kadastral veriler ve halihazır haritalar üzerinden seçilen 23 adet yer kontrol noktası, PAN için ise 12 yer kontrol noktası da kullanılmıştır. Düzeltme sonrasında piksellere atanacak yeni radyometrik değerler için en yakın mesafe yöntemi kullanılmıştır. Yapılan geometrik düzeltme sonucunda ortaya çıkan hata miktarları MS görüntü X yönünde 1.1990 m, Y yönünde 0.9897m olmak üzere toplam RMS değeri, 1.5547m ve PAN için da X yönünde 0.2287, Y yönünde 0.2297 ve toplam RMS değeri, 0.3241m olarak elde edilmiştir. Kış dönemi görüntüleri ise bu geometrik düzeltmesi yapılan görüntülere göre düzeltilmiştir.

2. YÖNTEM

2.1 Rastgele Orman Sınıflandırma Yöntemi

İlk olarak sınıflandırma işleminde kullanılacak sınıfların daha sağlıklı seçilebilmesi için yer doğrulama yapılmış ve çalışma bölgesine gidilerek El Tipi GPS yardımıyla fındık ve çay sınıfları için örnek alanlar toplanmıştır. Bu tarımsal ürünler çalışma bölgesinin farklı bölgelerinde farklı spektral özellik gösterebilecekleri için farklı alanlardan veriler toplanmıştır. Bu toplanan örnek alanlar göz önünde bulundurularak MS uydu görüntüsünden 7 farklı sınıf (orman, fındık, çay, toprak, gölge, kentsel alan1 ve kentsel alan2) için örnek pikseller toplanmıştır. Yaz dönemi görüntüsündeki sınıflar için toplamda 1755 piksel, kış dönemi görüntüsü için de 1779 piksel toplanmıştır. Her sınıf için yaklaşık olarak eşit sayıda eğitim pikseli kullanılmıştır. Daha sonraki aşamada RO sınıflandırma yönteminde kullanılmak üzere eğitim ve test verileri oluşturulmuştur. Eğitim ve test verilerinin oluşturulması ve sınıflandırma işlemlerinde Matlab programında gerçekleştirilmiştir.

Eğitim veri setleri kullanılarak farklı N ve m kombinasyonları için elde edilen OOB hatası, test doğruluğu, kapa ve hesaplama zamanı değerleri Tablo1 ve Tablo2'de verilmiştir. Tablo1'de görüldüğü gibi, yaz görüntüsüne ait WorldView-II (MS) uydu görüntüsü için en optimum sonucu N= 300 ve m=3 parametreleri vermiştir. Kış görüntüsüne ait WorldView- II (MS) uydu görüntüsü için ise N= 350 ve m=3 parametreleri seçilmiştir (Tablo2).

Tablo 1.Yaz görüntüsüne ait çalışma alanı için RO parametrelerinin testi

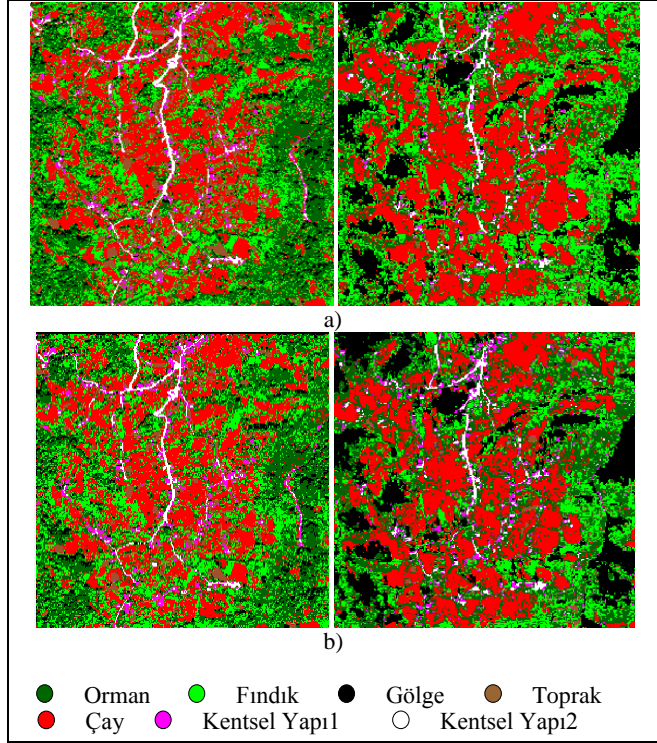
N	m	OOB hatası (%)	Test Doğruluğu (%)	Kappa	Hesaplama zamanı (sn)
1265	3	0.071795	92.82	0.9271	279.6308
1000	3	0.071795	92.82	0.9266	231.3488
833	3	0.066667	93.33	0.9320	189.2905
300	3	0.056410	94.36	0.9423	80.9762

Tablo 2.Kış görüntüsüne ait çalışma alanı için RO parametrelerinin testi

N	m	OOB hatası (%)	Test Doğruluğu (%)	Kappa	Hesaplama zamanı (sn)
1265	3	0.077572	92.24	0.9203	809.5525
1000	3	0.067454	93.25	0.9315	210.8681
833	3	0.067454	93.25	0.9312	180.1689
350	3	0.064081	93.59	0.9347	83.3812

Öncelikle sadece spektral değerler kullanılarak uygun parametrelerle sınıflandırma işlemi gerçekleştirilmiştir. Sonraki aşamada EDM ile doku çıkarımı işlemi yapılmış. Bu yöntem ile doku çıkarımı işlemi için ENVI programından yararlanılmıştır.

3x3, 5x5, 7x7 ve 9x9 gibi farklı filtre boyutlarındaki pencere matrisleri (1,1) çapraz yönde, (1,0) yatay yönde ve (0,1) düşey yönde olmak üzere farklı yönlerde uygulanmıştır. Gri değer seviyesi 64 olarak seçilmiştir. Üretilen EDM' den kontrast, farklılık (dissimilarity), homojenlik, enerji, entropi, ortalama, varyans ve korelasyon gibi farklı istatistiksel özellik hesaplatılmıştır. Bu doku çıkarımı işlemi hem yaz görüntüsüne hem de kış görüntüsüne uygulanmıştır. Ve bu özellikler de RO sınıflandırma işlemine katılarak bu özelliklere göre de görüntüler sınıflandırılmıştır.



Şekil 4. a)Yaz (solda) ve Kış (Sağda) dönemlerine ait sadece spektral değerler kullanılarak elde edilen RO görüntü b) Yaz (solda) ve Kış (Sağda) dönemlerine ait EDM doku özelliklerinin RO ile sınıflandırılması

Sınıflandırma sonuçlarının değerlendirilmesinde Congalton ve Green (1999)'ün önerdiği multi nominal dağılım yaklaşımı kullanılmıştır. 735 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 genel doğruluğu aynı 735 nokta kullanılarak test edilmiştir. Her bir sınıflandırma sonucunun doğruluğu, en yaygın kullanılan sınıflandırma sonrası doğruluk analizleri yöntemlerinden biri olan hata matrisi kullanılarak değerlendirilmiştir.

Tablo 3'de yaz görüntüsüne ait sınıflandırma sonuçlarından elde edilen hata matrisleri incelendiğinde genel olarak tüm yöntemlerde benzer spektral özelliğe sahip olan sınıfların birbiriyle karıştığı görülmektedir. Yaz dönemine ait WorldView-II görüntüsünde yeşil tonlarının hakim olduğu Orman, fındık ve çay sınıfları benzer spektral özelliğe sahip oldukları için en çok karışan sınıflar olmuşlardır. Ayrıca ağaçların gölgeleri görüntü üzerinde farklı tonlarda görülmektedir. Bazı kısımlarda koyu yeşil bazı bölgeler de ise siyah olarak görüldüğünden gölge sınıfında da karışmalar görülmüştür. Tablo3.a' daki hata matrisinde göre görüldüğü gibi sadece spektral değerler kullanıldığında toplam 735 noktadan

581'i kendi sınıfında temsil edilmiş ve RO' nun genel doğruluğu %79.05 olarak belirlenmiştir.

Tablo 3b'deki hata matrisi incelendiğinde 735 rastgele atılmış noktadan 586'ı kendi sınıfında sınıflandırılmıştır ve RO için %84.08 genel doğruluk sağlamıştır. Tablo 3.a ve 3.b karşılaştırıldığında yaz dönemi için fındık sınıfının üretici ve kullanıcı doğruluklarında sırasıyla %11 ve %14, çay sınıfının üretici doğruluğunda ise %14 oranlarında iyileşme görülmüştür. Çay sınıfının kullanıcı doğruluğuna bakıldığında ise sadece spektral değerlerle yapılan RO, EDM' den %1 oranında çayı daha başarılı belirlemiştir. Sadece spektral değerlerle yapılan RO' da 735 noktadan 83 adet tanımlanan fındık örneğinden 51 tanesi fındık olarak sınıflandırılırken EDM' de bu üründen 63 tanesi kendi sınıfında temsil edilmiştir. Çay için ise sadece spektral değerlerle yapılan RO' da 735 noktadan 139 adet tanımlanan çay örneğinden 96 tanesi çay olarak sınıflandırılırken EDM' de bu üründen 116 tanesi kendi sınıfında temsil edilmiştir.

Tablo 3. Yaz görüntüsü için a) spektral değerlere ait RO sınıflandırma sonucu hata matrisleri b) EDM doku özelliklerinin sınıflandırma sonucu hata matrisleri

		Referans								Üretici Doğ.(%)	Kullanıcı Doğ.(%)
		Orman	Fındık	Gölge	Toprak	Çay	Kentsel Y.1	Kentsel Y.2	ΣSatur		
Sınıflar	Orman	76	17	10	0	1	1	0	105	74.51	72.38
	Fındık	16	51	0	2	36	0	0	105	61.45	48.57
	Gölge	4	1	99	0	1	0	0	105	90.83	94.29
	Toprak	3	7	0	75	5	13	2	105	89.29	71.43
	Çay	2	7	0	0	96	0	0	105	69.06	91.43
	Kentsel Y.1	1	0	0	7	0	88	9	105	79.28	83.81
	Kentsel Y.2	0	0	0	0	0	9	96	105	89.72	91.43
	ΣSütun	102	83	109	84	139	111	107	735		
		Genel Doğruluk= 79.05% Kappa= 0.7537									

		Referans								Üretici Doğ.(%)	Kullanıcı Doğ.(%)
		Orman	Fındık	Gölge	Toprak	Çay	Kentsel Y.1	Kentsel Y.2	ΣSatur		
Sınıflar	Orman	79	12	3	2	1	1	0	98	77.45	80.61
	Fındık	17	63	2	2	22	0	0	106	75.90	59.43
	Gölge	2	0	104	1	0	0	0	107	95.41	97.20
	Toprak	1	1	0	69	0	4	2	77	82.14	89.61
	Çay	3	7	0	1	116	0	1	128	83.45	90.63
	Kentsel Y.1	0	0	0	8	0	101	18	127	90.99	79.53
	Kentsel Y.2	0	0	0	1	0	5	86	92	80.37	93.48
	ΣSütun	102	83	109	84	139	111	107	735		
		Genel Doğruluk= 84.08% Kappa= 0.8137									

Kış görüntülerinde genel olarak yaz görüntüsünde olduğu gibi spektral özellikleri benzer olan sınıfların birbiriyle karıştığı ancak bu durumun yaz görüntüsünden daha az olduğu görülmüştür. Çünkü kış görüntüsünde fındık ağaçlarının yaprak dökmesi bazı ağaçların sararması ve çay alanlarının yeşilliklerini koruması sebebiyle spektral benzerlikler azalmış ve farklılıklar ortaya çıkmıştır. Her ürün kendi içerisinde değerlendirildiğinde aynı ürünün farklı sezonlarda daha başarılı temsil edildiği ve bu ürünlerin birbirinden daha iyi ayrıldığı görülmüştür. Ancak genel kış döneminde güneş ışınlarının

daha eğik gelmesi sebebiyle eğimin yüksek olduğu bölgelerde gölgeler oluşmuş ve o bölgede yer alan ürünler tespit edilememiştir. Bu da genel doğruluğu düşürmüştür (Tablo 4).

Tablo 4. Kış görüntüsü için a) spektral değerlere ait RO sınıflandırma sonucu hata matrisleri b) EDM doku özelliklerinin sınıflandırma sonucu hata matrisleri

		Referans								Üretici Doğ.(%)	Kullanıcı Doğ.(%)
		Orman	Fındık	Gölge	Toprak	Çay	Kentsel Y.1	Kentsel Y.2	ΣSatur		
Sınıflar	Sınıfsız	1	1						2
	Orman	88	24	1	28	7	20	9	177	62.41	49.72
	Fındık	27	75	4	7	4	4	1	122	68.81	61.48
	Gölge	6	1	107	0	1	1	6	122	93.04	87.70
	Toprak	5	1	0	33	0	1	4	44	42.86	75.00
	Çay	15	4	2	2	107	0	1	131	89.92	81.68
	Kentsel Y.1	0	0	0	5	0	46	4	55	59.74	83.64
	Kentsel Y.2	0	3	0	2	0	5	72	82	74.23	87.80
ΣSütun	141	109	115	77	119	77	97	735			
		Genel Doğruluk= 71.84% Kappa= 0.6669									

		Referans								Üretici Doğ.(%)	Kullanıcı Doğ.(%)
		Orman	Fındık	Gölge	Toprak	Çay	Kentsel Y.1	Kentsel Y.2	ΣSatur		
Sınıflar	Sınıfsız	1	0	0	0	0	0	0	1
	Orman	83	14	4	8	12	2	3	126
	Fındık	30	68	7	4	3	0	0	112	58.87%	65.87%
	Gölge	0	2	104	0	2	0	0	108	62.39%	60.71%
	Toprak	12	16	0	58	1	2	2	91	90.43%	96.30%
	Çay	5	0	0	0	100	0	0	105	75.32%	63.74%
	Kentsel Y.1	7	2	0	6	1	69	5	90	84.03%	95.24%
	Kentsel Y.2	3	7	0	1	0	4	87	102	89.61%	76.67%
ΣSütun	141	109	115	77	119	77	97	735			
		Genel Doğruluk= 77.41% Kappa= 0.7356									

3. SONUÇLAR

Yapılan çalışmada ürünlerin sadece MS uydu görüntüsünden alınan spektral değerleri dikkate alınarak RO sınıflandırıcı ile sınıflandırmaları sonucu %79.05 genel sınıflandırma doğruluğu elde edilmiştir. Daha sonraki aşamada PAN görüntüsü üzerinde farklı filtre boyutlarının farklı yönlerde uygulanması ile EDM oluşturulmuş ve bu matris yardımıyla kontrast, farklılık, homojenlik, enerji, entropi, ortalama, varyans ve korelasyon gibi farklı istatistiksel özellik hesaplanmış doku özellikleri çıkarılmıştır. Spektral özelliklerin yanında bu istatistiksel doku özellikleri kullanılarak ürünler RO ile sınıflandırılmış ve %84.08 genel sınıflandırma doğruluğu elde edilmiştir. Elde edilen tüm sonuçlar Tablo 4'te verilmiştir. Buna göre doku özelliğinin kullanımı RO' nun sınıflandırma doğruluğunu %5 oranında artırdığı gözlenmiştir. Ayrıca ürünlerden elde edilen doğruluklar da irdelendiğinde fındık alanlarında ortalama %13, çay alanlarında ise ortalama %7 oranında artış görülmüştür. Elde edilen sonuçlar doku özelliğinin sınıflandırmadaki başarısını ortaya koymuştur.

Teşekkür

Bu çalışmada kullanılan veriler “Doğu Karadeniz Bölgesindeki Tarımsal Ürünlerin Uzaktan Algılama Teknolojileriyle Tespiti ve Analizi” başlıklı, 111Y296 nolu TÜBİTAK 1001 projesi tarafından temin edilmiştir. TÜBİTAK' a desteğinden dolayı teşekkür ederiz.

Kaynaklar:

Aksoy, S., 2009. Automatic Detection of Hedges and Orchards Using Very High Spatial Resolution Imagery, in 15th GeoCAP Conference on Geomatics in Support of the EDMon Agricultural Policy, Taormina, Italy, November 18-20.

Archer, K. J., 2008. Empirical characterization of random forest variable importance measure, *EDMputational Statistics & Data Analysis*, 52(4), 2249-2260.

Breiman, L. ve 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.

Congalton, R. G. ve Green, K. 1999. Assessing the accuracy of remotely sensed data: principles and practices. Boca Raton: Lewis Publishers.

Cutler, D.R., Edwards, T.C. Jr, Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. ve Lawler, J.J., 2007. Random forests for classification in ecology, *Ecology*, 88, 2783–2792.

Freund, Y. ve 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, ve 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 EDMputing (DSC 2003)*, Vienna, Austria, URL: <http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/FurlanelloEtAl.pdf> (20.11.2010).

Gislason, P.O., Benediktsson, J.A ve Sveinsson, J.R., 2006. Random forest For Land Cover Classification, *Pattern Recognition Letters*, 27(2006), 294-300.

Jay, S., Lawrence, R., Repasky ve K., Keith, C., 2009. Invasive Species Mapping Using Low Cost Hyperspectral Imagery, *ASPRS 2009 Annual Conference Baltimore*, Maryland ♦ March 9-13, 2009.

Liaw A. ve Wiener M., 2002. Classification And Regression By Random Forest, *R News*, Vol.2/3, December.

Ntouros, K. D., Gitas, I. Z. ve Silleos, G. N., 2009. Mapping Agricultural Crops With Eo-L Hyperion Data, 978-1-4244-4687-2/09/\$25.00 ©2009 IEEE.

Mather, P. M., 2004. *EDMputer Processing of Remotely-Sensed Images: An Introduction*, Third edition, Wiley, USA, ISBN 0-470-84918-5.

Ok, Asli O., Akar, O. ve Gungor, O., 2012. Evaluation of random forest method for agricultural crop classification, *European Journal of Remote Sensing*, Vol. 45 Issue 3, p421.

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.

Richards, J.A. ve Jia, X., 2006. *Remote Sensing Digital Image Analysis: An Introduction 4th Edition*, Springer-Verlag, Germany, ISBN-10: 3-540-25128-6.

Takçı, H., 2008. *Veri Madenciliği Dersi Notları*, (Veri Madenciliği Dersi – GYTE – Dr. Hidayet Takçı 10/05/2008)

Tso, B. ve Mather P. M., 2009. *Classification Methods For Remotely Sensed Data*, Second Editon, Taylor & Francis Group, United States of America.

Watts, J. D., Powell, S.L., Lawrence, R. L. ve 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. ve 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.

Yalınız, I.Z. ve Aksoy, S., 2009. Uydu Görüntülerinde Düzenli Dikim Alanlarının Belirlenmesi (in Turkish)," in 17. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı, Antalya, Turkey, April 9-11.