

Kanonik Korelasyon Orman Algoritması ile Uzaktan Algılanmış Görüntülerin Sınıflandırılması

İsmailÇölkesen¹, Taşkın Kavzoğlu²

^{1,2}Gebze Teknik Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, Gebze
e-posta:icolkesen@gtu.edu.tr

Geliş Tarihi: 01.02.2017 ; Kabul Tarihi: 02.08.2017

Özet

Doğru ve güvenilir tematik bilgiye sahip olmak küresel ve yerel ölçekli birçok çalışma için esastır. Uydu görüntülerinin sınıflandırılmasıyla yeryüzünün farklı fiziksel özelliklerini temsil eden tematik haritaların üretilmesi uzaktan algılamada en yoğun çalışma konularından birisidir. Bu amaçla günümüze kadar birçok sınıflandırma algoritması geliştirilmiş ve uygulanmıştır. Sınıflandırıcıların birleşimi veya çoklu sınıflandırma sistemleri olarak bilinen toplu öğrenme algoritmaları karmaşık ve gürültülü veri setleri kullanımındaki başarıları nedeniyle birçok sınıflandırma işleminde yaygın olarak kullanılmaktadır. Kanonik korelasyon ormanı (KKO) algoritması, her bir karar ağacının kanonik korelasyon bileşenleri kullanılarak oluşturulduğu yeni nesil bir karar ormanı olarak son yıllarda önerilen bir toplu öğrenme algoritmasıdır. Bu çalışmada, Temmuz 2016 tarihinde elde edilen multispektral Landsat OLI görüntüsü kullanılarak çalışma alanının arazi kullanımı ve arazi örtüsü haritası üretiminde KKO algoritması kullanılmıştır. Yöntemin sınıflandırma performansı popüler en çok benzerlik (EÇB) ve karar ağaçları (KA) algoritmaları ile karşılaştırılmış ve istatistiksel olarak analiz edilmiştir. Sonuçlar, KKO algoritmasının genel doğruluk açısından EÇB ve KA algoritmalarına kıyasla üstün performans gösterdiğini ortaya koymaktadır. Genel sınıflandırma doğruluğundaki iyileşmenin %11'e ulaştığı ve sınıflandırma performansındaki bu iyileşme düzeyinin istatistiksel olarak anlamlı olduğu görülmüştür. Sonuç olarak, orman benzeri toplu öğrenme algoritmalarının yeni bir üyesi olarak önerilen KKO algoritmasının, multispektral uydu görüntüsünün sınıflandırılmasında etkili bir sınıflandırıcı olduğu görülmüştür.

Anahtar kelimeler

Sınıflandırma; Toplu öğrenme; Kanonik korelasyon ormanı; Karar ağaçları; En çok Benzerlik.

Classification of Remotely Sensed Imagery with Canonical Correlation Forest Algorithm

Abstract

Having accurate and reliable thematic information is essential for many global and local scale studies. Producing thematic maps representing different physical characteristics of the Earth's surface by means of image classification has been one of the most concentrated issues in remote sensing. Up to now, many classification algorithms have been proposed and applied for this purpose. Ensemble learning algorithms, also known classifier ensemble or multiple classifier systems, have been widely-used in various supervised classification tasks due to their robustness in handling complex and noisy datasets. An ensemble learning algorithm called canonical correlation forest (CCF) has been recently introduced as a new type of decision forest where each individual tree is constructed using canonical correlation components. In this study, CCFs classifier was used to produce land use and land cover map of the study area using multispectral Landsat OLI image acquired in July 2016. The classifier performances were compared and statistically analysed with popular maximum likelihood (ML) and decision tree (DT) classifiers. Results showed that CCF algorithm showed superior performance compared to ML and DT in terms of overall accuracy. The improvement in overall classification accuracy reached to 11% and this level of improvement in classification performance was found to be statistically significant. To sum up, the CCF algorithm, proposed as a new member of the forest-like ensemble learning algorithms, was found as an effective classifier for classifying multispectral satellite images.

Keywords

Classification; Ensemble learning; Canonical correlation forest; Decision trees; Maximum likelihood

© Afyon Kocatepe Üniversitesi

1. Giriş

Uydu görüntülerinin sınıflandırılması yardımıyla arazi örtüsü ve arazi kullanımına ilişkin bilgilerin elde edilmesi uzaktan algılama alanında en önemli çalışma konularından birisidir. Sınıflandırma sonucu elde edilen tematik haritalar çevresel modelleme, değişim analizi, tarımsal içerikli araştırmalar ve arazi kullanımı planlamaları gibi birçok uygulamada temel altlık olarak kullanılmaktadır. Bu nedenle tematik haritaların doğruluğu ve yeryüzü özelliklerinin temsiliyeti noktasındaki güvenilirliği söz konusu çalışmalar için büyük önem arz etmektedir. Sınıflandırma işleminin başarısı tematik harita doğruluğu ile doğrudan ilişki olup; ilgi duyulan alandaki yeryüzü özellikleri, uzaktan algılanmış görüntülerin çözünürlüğü ve uygun bir sınıflandırma yönteminin seçimini gibi çeşitli faktöre bağlıdır (Lu and Weng 2007). Uydu görüntüleri yardımıyla doğru ve güvenilir tematik haritalar üretilmesi amacıyla günümüze kadar pek çok sınıflandırma tekniği geliştirilmiştir (Lu and Weng 2007; Tso and Mather 2009; Li *et al.* 2014). Özellikle son yıllarda, klasik parametrik sınıflandırma algoritmaları yerine karmaşık sınıflandırma probleminin modellenmesi için istatistiksel kabullere dayanmayan sınıflandırma algoritmalarının kullanımı artış göstermiştir. Bu algoritmalar arasında destek vektör makineleri, karar ağaçları, hibrit sınıflandırma modelleri (örneğin, fonksiyonel model ağaçlar ve lojistik model ağaçlar) ve toplu öğrenme algoritmaları (örneğin, rastgele orman ve rotasyon orman) sınıflandırma doğruluğundaki başarıları nedeniyle literatürde yaygın olarak kullanılmaktadır (Kavzoglu and Mather 2003; Belgiu and Drăguț 2016; Colkesen and Kavzoglu 2017; Çölkesen, 2017).

Sınıflandırma algoritmalarının birleştirilmesi veya kombinasyonu olarak da bilinen toplu öğrenme algoritmalarının sınıflandırma ve regresyon problemlerinin çözümünde kullanımı makine öğrenmesi veya veri madenciliği alanında önemli bir araştırma konusudur (Dietterich 2000; Briem *et al.* 2002; Rokach 2010; Jurek *et al.* 2014). Toplu öğrenme algoritmalarının temel çalışma prensibi, birden fazla sınıflandırıcının bir arada kullanımı ve

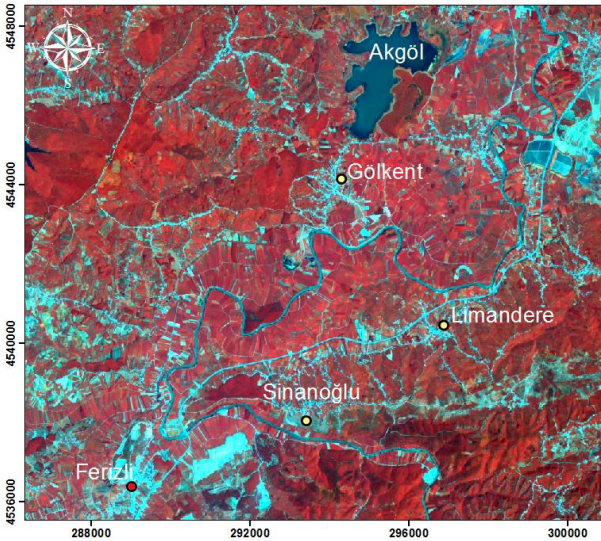
her bir sınıflandırıcının tahminlerinin birleşimi neticesinde sınıflandırma işleminin gerçekleştirilmesidir (Dietterich 2000). Bu yaklaşımla tek bir sınıflandırma algoritması kullanımına göre daha yüksek sınıflandırma doğruluklarının elde edildiği birçok çalışmada ifade edilmiştir (Opitz and Maclin 1997; Gislason *et al.* 2006; Kavzoglu and Colkesen 2013). Birden çok karar ağacı sınıflandırıcısının bir arada kullanılması esasına dayanan ve karar ormanı olarak da isimlendirilen toplu öğrenme algoritmaları farklı karakteristiklere sahip uydu görüntülerinin sınıflandırmasında başarıyla kullanılmaktadır (Pal, 2005; Kavzoglu and Colkesen 2013; Kavzoglu *et al.* 2015). Karar ağaçlarını esas alan toplu öğrenme algoritmalarının çalışma prensibi orman yapısını oluşturan her bir karar ağacının farklı bir veri seti kullanılarak eğitilmesi ve sınıflandırma sonuçlarının birleşimi neticesinde sonuç tahminin yapılmasıdır. Bu açıdan ele alındığında toplu öğrenme algoritmasının başarısı farklı eğitim verilerinin oluşturulması veya karar ormanı içerisindeki çeşitliliğin sağlanmasıyla doğrudan ilişkilidir (Diez-Pastor *et al.* 2015; Ren *et al.* 2016). Karar ormanındaki çeşitliliğin artırılması ve dolayısıyla sınıflandırma doğruluğunun artırılmasına yönelik olarak geliştirilen kanonik korelasyon ormanı (KKO) yeni nesil bir toplu öğrenme algoritmasıdır. Bu toplu öğrenme modelinde ağaç yapıları arasındaki değişkenliğin artırılması amacıyla, kanonik korelasyon analizi yardımıyla hesaplanan bileşenler kullanılarak karar ağaçlarının oluşturulması söz konusudur (Rainforth and Wood 2015).

Bu çalışmada, KKO algoritmasıyla 2016 tarihli Landsat OLI görüntüsü yardımıyla, Sakarya ili Ferizli ilçesini içerisine alan ve tarımsal arazilerin çoğunlukta olduğu çalışma alanına ilişkin tematik harita üretilmesi amaçlanmıştır. Yöntemin sınıflandırma performansının karşılaştırılması amacıyla en çok benzerlik (EÇB) ve karar ağaçları (KA) algoritmaları kullanılarak tematik harita üretimi gerçekleştirilmiş ve sonuçlar Wilcoxon işaretli sıralar testi kullanılarak istatistiksel olarak analiz edilmiştir. Sonuçların karşılaştırılmasında, genel sınıflandırma doğruluğu, Kappa değeri,

kullanıcı ve üretici doğruluklarından yararlanılmıştır.

2. Çalışma Alanı ve Kullanılan Veri

Bu çalışmada Sakarya ilinin kuzeyinde yer alan Ferizli ilçe sınırları içerisindeki Sakarya nehrinin çevrelediği Sinanoğlu, Limandere ve Akgöl mahallelerini kapsayan bölge çalışma alanı olarak belirlenmiştir (Şekil 1). Çalışma alanı ekonomik değeri olan başta mısır, fındık ve şekerpancarı olmak üzere birçok tarımsal ürün çeşidinin yetiştirildiği yaklaşık 2.000 hektarlık bir alanı kapsamaktadır. Bölgede temel bitkisel üretim mısır ve fındık üzerine yoğunlaşmıştır. 2015 tarımsal üretim istatistiklerine göre Sakarya ili fındık üretiminde dördüncü, mısır üretiminde ise sekizinci sırada yer almaktadır (www.tuik.gov.tr). Bununla birlikte çalışma sahası içerisinde şekerpancarı, yonca ve çeşitli bahçe bitkileri de (domates, karpuz, fasulye, biber vb.) yetiştirilmektedir.



Şekil 1. Çalışma alanı konumu ve kullanılan uydu görüntüsü.

Bu çalışmada sınıflandırmaya esas veri seti olarak multispektral Landsat OLI uydu görüntüsü kullanılmıştır. Mevcut hâlihazır haritalar ve bölgeye ait yüksek çözünürlüklü görüntüler kullanılarak, Landsat görüntüsünün geometrik düzeltilmesi ve UTM koordinat sistemine dönüşümü gerçekleştirilmiştir. Sınıflandırma işlemi öncesinde multispektral bantların mekânsal çözünürlüğünün

arttırılması amacıyla Landsat görüntüsünün multispektral ve pankromatik bantları kaynaştırılmıştır. Kaynaştırma işleminin gerçekleştirilmesi amacıyla literatürde farklı algoritmaların kullanımı söz konusudur. Fakat, kaynaştırma algoritmasının seçimi kullanılan uydu görüntüsü ve çalışma alanına göre değişkenlik göstermektedir (Ghosh and Joshi 2014). Bu çalışmada, kaynaştırma işleminin gerçekleştirilmesi amacıyla Brovey, temel bileşenler analizi (PCA) ve Gram-Schmidt (GS) algoritmaları değerlendirmeye alınmıştır. Algoritmaların kaynaştırma işlemindeki performanslarının değerlendirilmesinde yaygın olarak kullanılan entropi ölçütünden yararlanılmıştır. Entropi ölçütüne ilişkin sonuçlar ve literatürde son yıllarda aynı sensör kullanılarak gerçekleştirilen çalışmalar birlikte değerlendirilerek kaynaştırma işlemi için GS algoritmasının kullanılmasına karar verilmiştir (Huang *et al.* 2016; Zhanget *al.* 2016). Kaynaştırma işleminin gerçekleştirilmesinde 3x3 görüntü yumuşatma filtresi uygulanmış ve kaynaştırma işlemi sonrasında yeniden örnekleme işlemi en yakın komşuluk algoritmasıyla gerçekleştirilmiştir.

3. Sınıflandırma Algoritmaları

Bu çalışmada, uydu görüntüsünün sınıflandırılması ve tematik harita üretiminde kanonik korelasyon ormanı, karar ağaçları ve en çok benzerlik sınıflandırıcısı kullanılmıştır.

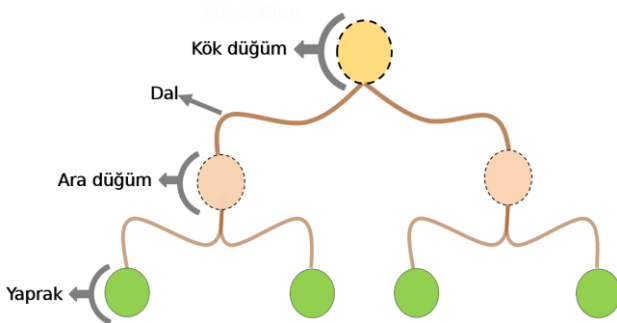
3.1. En çok benzerlik algoritması

En çok benzerlik (EÇB) yöntemi, uzaktan algılanmış görüntülerin sınıflandırılması ve tematik harita üretiminde en sık kullanılan bir kontrollü sınıflandırma yöntemidir. EÇB yöntemi sınıflandırma işleminde ortalama değer, varyans ve kovaryans gibi istatistiki bilgileri dikkate aldığından istatistiksel tabanlı sınıflandırma algoritması olarak da bilinmektedir. Yöntemin temel çalışma prensibi, eğitim veri setini oluşturan sınıflara ilişkin olasılık fonksiyonlarının hesaplanması ve herhangi bir pikselin hesaplanan olasılıklara göre hangi sınıfa ait olduğunun tespit edilmesi esasına dayanmaktadır. İstatistiksel bir yöntem olan EÇB anlaşılması kolay

ve hızlı bir yöntem olmasına karşın, eğitim verilerinin ve sınıfları oluşturan piksellerin normal dağılımda olduğu kabulü nedeniyle karmaşık yapıdaki sınıflandırma problemlerinin çözümünde yetersiz kalabilmektedir (Mather 1999).

3.2. Karar ağaçları algoritması

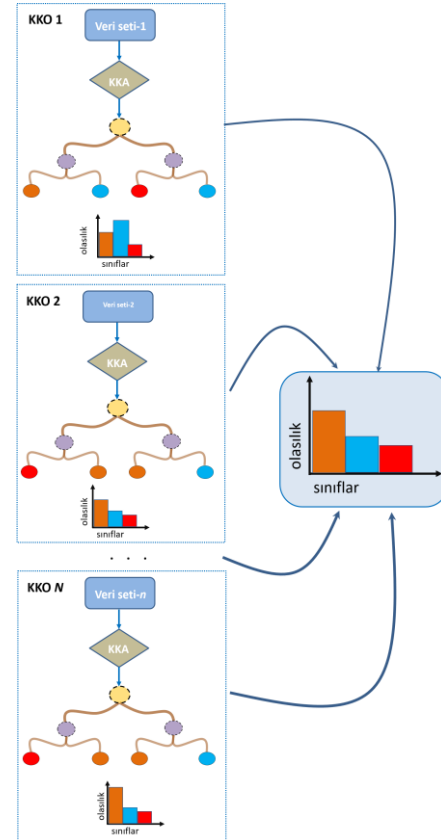
Karar ağaçları (KA) uzaktan algılanmış görüntülerin sınıflandırılmasını da içerisine alan birçok örüntü tanıma problemde yaygın olarak kullanılan kontrollü sınıflandırma yöntemidir. KA'nın temel çalışma prensibi, herhangi bir sınıflandırma probleminin çok aşamalı bir hale getirilerek basit bir karar verme işleminin gerçekleştirilmesi esasına dayanmaktadır (Safavian and Landgrebe 1991; Kavzoğlu ve Çölkesen 2010). Akış şemasına benzeyen bir yapıya sahip olan karar ağaçları Şekil 2'den de görüleceği üzere düğüm, dal ve yaprak olarak ifade edilen üç temel kısımdan oluşmaktadır. Ağaç yapısında her bir spektral bant (öznitelik) bir düğüm noktası ile temsil edilirken, ağacın en üst noktası veya ilk düğüm noktası kök düğüm olarak ifade edilmektedir. Ağaç yapısının en alt kısmında, ulaşılmak istenen hedef sınıfları temsil eden yapraklar bulunmaktadır. Basit bir ağaç yapısında düğüm noktaları ile yapraklar arasındaki ilişki dallar aracılığıyla sağlanmaktadır. Sınıflandırma işleminde öncelikli olarak eğitim verilerinden yararlanılarak ağacın dallanmasına esas teşkil edecek karar kuralları ve sınıflandırma modeli oluşturulur. Herhangi bir pikselin sınıflandırılmasında ise, ağacın kök düğümünde sorular sorulmaya başlanır ve bir sınıfı temsil eden yaprağa ulaşılan kadar işlem devam eder (Quinlan 1993; Pal and Mather 2003; Kavzoğlu ve Çölkesen 2010).



Şekil 2. Basit bir karar ağacı algoritması yapısı.

3.3. Kanonik korelasyon orman algoritması

Kanonik korelasyon ormanı (KKO) karar ağaçlarını esas alan ve kontrollü sınıflandırma işleminin gerçekleştirilmesi için tasarlanan yeni nesil bir toplu öğrenme algoritmasıdır (Rainforth and Wood 2015, Xia et al. 2017). Karar ağaçlarını esas alan bir toplu öğrenme algoritması olduğundan rastgele orman ve rotasyon orman algoritmalarını da içerisine alan karar ormanı sınıflandırıcıları arasında yer almaktadır. Algoritmanın temel çalışma prensibi Şekil 3'den de görüleceği üzere bir dizi kanonik korelasyon ağacının bir araya geldiği bir karar ormanı oluşturmaktır. Bu karar ormanı modelindeki her bir ağaç kanonik korelasyon (KK) analizi sonucunda elde edilen bileşenler kullanılarak eğitilir. KK analizi iki veya daha fazla değişken arasındaki ilişkilerin incelenmesinde kullanılan bir dönüşüm tekniği olup, KKO algoritması oluşumunda eğitim veri setindeki sınıflar ile özellikler (bantlar) arasındaki korelasyonu maksimum olmasını sağlayan özellik dönüşümleri gerçekleştirilir.



Şekil 3. Kanonik korelasyon ormanı çalışma prensibi.

prensibi.

Dönüşüm sonrasında elde edilen bileşenler yardımıyla ağacın dallanması ve gelişimi işlemleri gerçekleştirilir. Klasik karar ağacı modeli sınıflar arasında eksen bazlı (axis aligned) bölümlenmeler ile karar sınırları tanımlayarak sınıflandırma işlemi gerçekleşirken, kanonik korelasyon ağacı KK dönüşümü ile oluşan özellik uzayında sınıflar arasında eğimli (oblique) karar sınırları tanımlayabilmektedir. Sahip olduğu bu özellikle KKO algoritması rastgele orman ve rotasyon orman algoritmalarının gelişmiş bir versiyonu olarak ifade edilmektedir. Toplu öğrenme modelindeki ağaç sayısı karar ormanı algoritmalarının tamamı için kullanıcı tarafından belirlenen bir parametre olmasına rağmen, söz konusu parametredeki değişimler KKO algoritmasının sınıflandırma performansında istatistiksel olarak anlamlı değişimler ortaya çıkarmadığı ifade edilmektedir (Rainforth and Wood 2015). Dolayısıyla, rastgele orman algoritmasının aksine, KKO algoritmasının çok daha az sayıda ağaç ile sonuca ulaşabilmektedir.

4. Uygulama

Çalışma alanı içerisinde 9 temel arazi örtüsü/kullanım sınıfı olduğu tespit edilmiştir. Orman, bozkır, su ve toprak sınıfları çalışma alanının temel arazi örtüsü sınıfları olarak tanımlanmıştır. Bölge tarımsal ürün çeşitliliği açısından ele alındığında mısır ve fındık ekili alanların yoğunlukta olması nedeniyle mısır ve fındık olarak iki farklı tarımsal üretim sınıfı tanımlanmıştır. Bununla birlikte, bölge içerisindeki diğer tarımsal ürünler (şeker pancarı, yonca vb.) tarım sınıfı olarak adlandırılmış ve tek bir sınıf altında değerlendirmeye alınmıştır. Bölge içerisindeki asfalt yollar ve yapılaşma alanları için ayrı ayrı yol ve bina sınıfları tanımlanmıştır.

Kontrollü sınıflandırma probleminde kullanılacak eğitim veri seti boyutu sınıflandırma doğruluğundandır etkileyen faktörlerden

biridir. Özellikle en çok benzerlik algoritması gibi istatistiksel tabanlı sınıflandırıcılar fazla sayıda eğitim örneğine ihtiyaç duymakta ve sınırlı sayıda eğitim veri setleri kullanımı sonucunda bu algoritmaların sınıflandırma doğruluğu önemli derecede azalabilmektedir (Foody *et al.* 2006; Çölkesen ve Kavzoğlu 2010). Diğer taraftan, destek vektör makineleri ve karar ağaçları gibi istatistiksel tabanlı olmayan sınıflandırıcılar sınırlı sayıda eğitim kullanımında EÇB yöntemine göre daha doğru sınıflandırma sonuçları üretebilmektedir. Bununla birlikte yüksek sayıda eğitim verisi kullanımı, söz konusu algoritmaların sınıflandırma performansında istatistiksel olarak anlamlı değişiklikler ortaya çıkarmamaktadır (Kavzoğlu and Colkesen 2012). Bu çalışmada, istatistiksel ve istatistiksel olmayan sınıflandırıcılar bir arada kullanıldığından sınırlı sayıda eğitim veri seti kullanılmasından ziyade eğitim veri seti boyutu yüksek tutularak veri seti boyutunun sınıflandırma doğruluğuna etkisine ilişkin önyargı giderilmiştir. Çalışma alanında 9 temel arazi örtüsü sınıfı için tespit edilen örnekleme alanlarını kapsayan görüntü rastgele örnekleme prensibinden hareketle örneklendirilerek eğitim ve test veri setleri oluşturulmuştur. Sınıflandırma modellerinin oluşumu ve doğruluk analizlerinde objektifliğin sağlanması amacıyla eğitim ve test veri setleri belirlenirken tüm sınıflar için eşit sayıda örnek seçilmiştir. Sonuç olarak, 9 arazi örtüsü sınıfı için eğitim verisi olarak toplam 4.050 piksel (her bir sınıf için 450 piksel) ve test verisi olarak toplam 4.500 piksel (her bir sınıf için 500 piksel) seçilmiştir. Veri setinin EÇB, KA ve KKO algoritmaları yardımıyla sınıflandırılması ve tematik harita üretilmesi işlemleri Matlab 2013a yazılımı aracılığıyla gerçekleştirilmiştir. En çok benzerlik (EÇB) yöntemiyle gerçekleştirilen sınıflandırma işlemi neticesinde hesaplanan genel doğruluk ve Kappa değeri Tablo 1'den de görüleceği üzere %80,62 ve 0,78'dir.

Tablo 1. En çok benzerlik (EÇB), karar ağaçları (KA) ve kanonik korelasyon ormanı (KKO) algoritmaları için hesaplanan sınıflandırma sonuçları.

Algoritma	Genel doğruluk (%)	Kappa değeri
EÇB	80,62	0,78
KA	80,18	0,77
KKO	91,35	0,90

Karar ağaçları (KA) ile sınıflandırma işleminin gerçekleştirilmesinde yaygın olarak kullanılan C4.5 algoritmasından yararlanılmıştır. Ağaç modeli oluşumunda dallanmaya esas olacak bantların (özniteliklerin) tespitinde kazanç oranı kriteri kullanılmıştır. Eğitim veri seti ile oluşturulan ağaç yapısının sadeleştirilmesi ve işlem karmaşıklığının azaltılması maksadıyla ön budama yöntemi kullanılmıştır. Eğitim veri seti kullanılarak oluşturulan ağaç modelinde 366 düğüm ve sınıf etiketlerini içeren 367 yaprak bulunmaktadır. Ağaç modelinin kök düğümünde dallanmaya esas olacak özellik olarak Landsat OLI'nin ikinci orta kızılötesi (SWIR-2) bandının seçildiği, kök düğümden sonraki ikinci düzey bölünmelerde ise yakın kızılötesi (NIR) ve kırmızı bantların yer aldığı görülmüştür. Oluşturulan KA modelinin performansı test veri seti için analiz edildiğinde, modelin test veri setinde bulunan arazi örtüsü ve kullanım sınıflarına ilişkin örnek piksellere ait sınıf etiketlerini %80,18 genel doğruluk ve 0,77 Kappa değeri ile sınıflandırdığı görülmektedir (Tablo 1).

Model oluşumunda kullanılan ağaç sayısı veya karar ormanındaki ağaç sayısı tüm toplu öğrenme algoritmalarında olduğu gibi kanonik korelasyon ormanı (KKO) algoritmasının da kullanıcı tarafından belirlenen parametresidir. Ancak literatürde KKO ile gerçekleştirilen çalışmalarda söz konusu parametrede değişimlerinin sınıflandırma doğruluğu üzerinde istatistiksel olarak anlamlı değişimler ortaya çıkarmadığı, çok az sayıda kanonik korelasyon ağacında oluşan toplu öğrenme modeli ile çok fazla sayıda karar ağacının kullanıldığı rastgele orman algoritmasından daha iyi sınıflandırma sonuçlarına ulaşıldığı ifade edilmektedir (Rainforth and Wood 2015; Xia *et al.* 2017). Bu önemli avantajı nedeniyle KKO algoritması kullanıcı tarafından parametre tespiti gerektirmeyen bir algoritma olarak sınıflandırılmaktadır. Bu çalışmada KKO algoritması

için Rainforth and Wood (2015) tarafından önerilen ve 200 ağaçtan oluşan toplu öğrenme modeli kullanılmıştır. Oluşan KKO modelinin geçerliliği test veri seti kullanılarak analiz edildiğinde, Tablo 1'den de görüleceği üzere hesaplanan genel doğruluk ve Kappa değerleri %91,35 ve 0,90'dır.

Uygulama kapsamında multispektral Landsat OLI görüntüsünün sınıflandırılması ve tematik harita üretiminde kullanımı değerlendirmeye alınan KKO algoritmasının sınıflandırma performansı geleneksel EÇB ve KA algoritmaları ile elde edilen sonuçlarla karşılaştırıldığında önemli sonuçlara ulaşılmıştır. Tablo 1'de verilen sınıflandırma sonuçları analiz edildiğinde, en yüksek sınıflandırma doğruluğunun %91,35 ile KKO algoritması yardımıyla elde edildiği görülmektedir. Diğer taraftan EÇB ve KA algoritmaları ile uygulamada kullanılan veri seti için benzer sınıflandırma sonuçlarına (%80,62 ve %80,18) ulaşılmıştır. Elde edilen sonuçlar yeni nesil toplu öğrenme algoritması olarak ifade edilen KKO algoritması kullanımı ile sınıflandırma doğruluğunda önemli düzeyde bir iyileşme söz konusu olmuştur. Elde edilen sonuçlar tek bir karar ağacı algoritması kullanımı yerine birden fazla ağaçtan oluşan toplu öğrenme algoritması kullanımının sınıflandırma doğruluğunda önemli derecelerde (yaklaşık %11) artış ortaya çıkardığını göstererek, literatürde daha önce gerçekleştirilen çalışmalarda bu konu hakkındaki bulguları destekler niteliktedir.

Tablo 2. EÇB, KA ve KKO için hesaplanan kullanıcı (KD) ve üretici (ÜD) doğrulukları.

Sınıflar	Sınıflandırma algoritmaları					
	EÇB		KA		KKO	
	KD	ÜD	KD	ÜD	KD	ÜD
Tarım	52.24	60.60	58.20	63.96	85.60	83.43
Mısır	75.21	72.20	79.20	74.30	88.00	90.72
Orman	97.40	97.40	97.00	95.10	99.20	99.40
Fındık	96.25	97.60	96.40	92.16	98.60	96.10
Bozkır	61.67	78.20	68.80	66.93	86.60	86.77
Yol	80.53	73.60	69.00	73.56	85.60	83.59
Toprak	86.97	73.40	81.20	81.04	91.80	93.10
Bina	87.81	77.80	76.00	76.46	88.80	90.06
Su	99.37	94.80	95.80	96.18	98.00	99.39

Genel sınıflandırma doğruluklarının karşılaştırılmasının yanında oluşturulan hata matrisleri yardımıyla arazi örtüsü ve kullanım sınıflarına ilişkin kullanıcı ve üretici doğrulukları da hesaplanmış ve sonuçlar Tablo 2’de verilmiştir. Tablodan da görüleceği üzere EÇB, KA ve KKO algoritmaları tarafından orman, fındık ve su sınıflarına ilişkin hesaplanan kullanıcı ve üretici doğruluklar %90’ın üzerindedir. Ancak bozkır, mısır ve bölgedeki diğer tarım ürünleri temsil eden piksellerin sınıflandırılmasında KKO algoritmasının daha başarılı olduğu görülmektedir. KKO algoritması ile bozkır ve mısır sınıfı için hesaplanan kullanıcı ve üretici doğruluklarında %9-25 arasında artış olduğu görülürken, tarım sınıfı için doğruluk seviyesindeki artış %20-30 düzeyindedir. Toprak, yol ve bina sınıflarını temsil eden piksellerin sınıflandırılmasında da KKO algoritması kullanımının kullanıcı ve üretici doğruluğu üzerindeki pozitif etkileri tablodan da görülmektedir.

Değerlendirmeye alınan algoritmaların sınıflandırma performansları arasındaki farklılıkların istatistiksel olarak anlamlılığının tespiti amacıyla Wilcoxon işaretli sıralar testinden yararlanılmıştır. İki sınıflandırma algoritmasının performans farklılıklarının sıralanması esasına dayanan istatistik test, hesaplanan farkların pozitif ve negatif farklar olarak ayrı ayrı sıralanmasını ve daha sonra birbirleriyle karşılaştırılmasını esas almaktadır. Hesaplanan test istatistiği, belirlenen güven aralığında kritik normal dağılım tablo değerinden büyük olduğunda değerlendirmeye alınan iki sınıflandırma sonucu arasında istatistik olarak anlamlı bir fark olduğu sonucuna varılır (Çölkesen 2017).

Tablo 3. Wilcoxon işaretli sıralar test sonuçları.

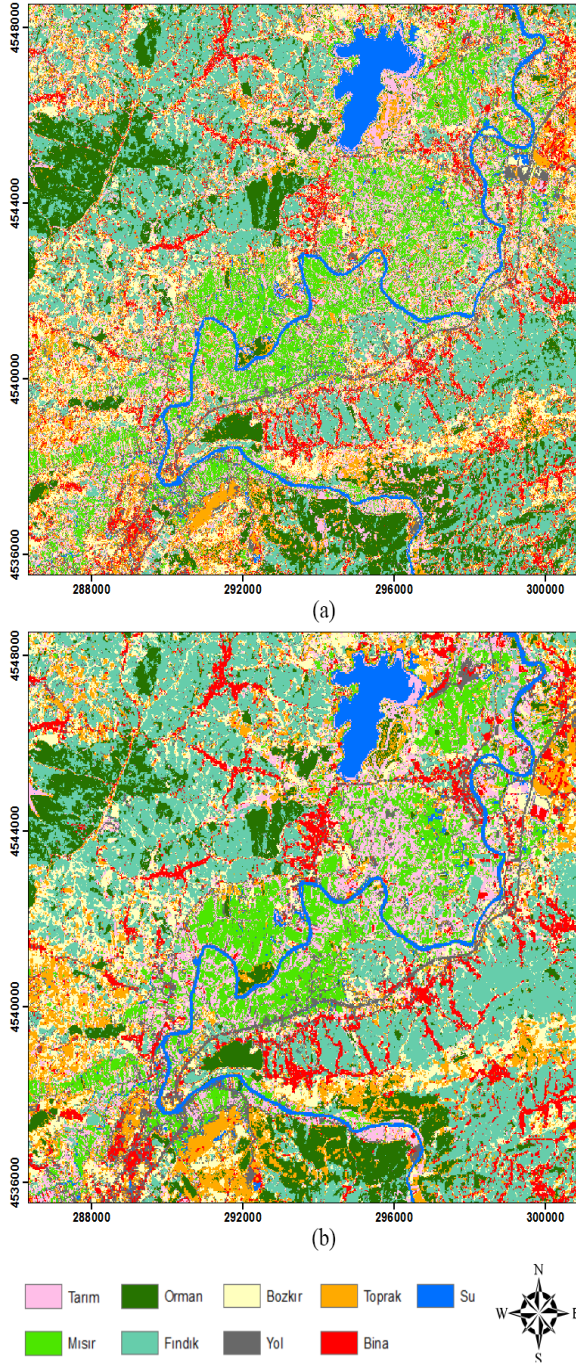
Karşılaştırma	Test istatistiği
EÇB - KA	0,71
KKO - EÇB	18,65
KKO - KA	19.36

Algoritmaların performansları ikili karşılaştırmalar şeklinde test edilmiş ve elde edilen istatistiksel

değerler Tablo 3’de verilmiştir. Tabloda %95 güven aralığında kritik tablo değerinden (1,96) küçük olan istatistiksel sonuçlar koyu olarak gösterilmiştir. Hesaplanan istatistiksel sonuçlardan da görüleceği üzere KKO ile elde edilen sınıflandırma sonucu ile EÇB ve KA algoritmalarından elde edilen sonuçlar arasındaki farklar %95 güven aralığında istatistiksel olarak anlamlıdır. Diğer bir ifadeyle, KKO algoritması kullanımı ile sınıflandırma doğruluğundaki %11’lik iyileşme istatistiksel olarak anlamlı bir artıştır. Diğer taraftan, test sonuçları EÇB ve KA algoritmalarının istatistiksel olarak benzer sınıflandırma performansı sergilediklerini göstermektedir. Aynı şekilde yeni bir toplu öğrenme algoritması olan KKO algoritmasının, tek bir karar ağacı kullanımına göre sınıflandırma sonuçlarında istatistiksel olarak anlamlı artış ortaya çıkardığı görülmektedir.

Uygulamada değerlendirmeye alınan tüm sınıflandırma algoritmalarına ait sınıflandırma modelleri uydu görüntüsüne uygulanarak çalışma alanına ilişkin tematik haritalar üretilmiştir. Söz konusu tematik haritalara örnek olarak KA ve KKO algoritmalarıyla üretilen tematik haritalar Şekil 4’de gösterilmiştir. KA ile üretilen tematik haritada (Şekil 4a) literatürde tuz-biber (salt-pepper) olarak bilinen gürültü oranının daha fazla olduğu ve dolayısıyla sınıflara ait doğal sınırların birbirinden ayırt edilmesinin güç olduğu görülmektedir. Özellikle çalışma bölgesi içerisinde canlı vejetasyonu gösteren mısır, fındık, orman ve diğer tarım alanlarını temsil eden pikseller arasındaki hatalı sınıflandırmalar neticesinde görüntü üzerinden ayırım yapmanın güçleştiği ifade edilebilir. Diğer taraftan KKO algoritması ile üretilen haritada (Şekil 4b) söz konusu problemlerden kaynaklı etkilerin daha az olduğu görülmektedir. Dolayısıyla KKO algoritması kullanımıyla sınıflandırma doğruluğundaki yaklaşık %11’lik iyileşmenin bir sonucu olarak haritada sınıflara ait sınıflar daha kolay tespit edilebilmektedir. Elde edilen bu sonuç özellikle benzer spektral özelliklere sahip fındık ve mısır gibi tarımsal ürün tiplerinin ayırt edilmesinde KKO algoritmasının daha başarılı olmasının doğal bir sonucu olarak ifade edilebilir.

Su geçirimsiz yüzeyler olarak bilinen kentsel alanlar ve asfalt yol sınıfına ait piksellere ilişkin sınıflandırma sonuçları incelendiğinde, KA ile üretilen tematik haritada (Şekil 4a) söz konusu piksellerin birbirine karıştığı ve görsel analizinin güç olduğu görülmektedir. Buna karşın KKO ile üretilen tematik haritada (Şekil 4b) şehir ve asfalt yollara ait piksellerin sınıflandırılmasının daha doğru olduğu ifade edilebilir.



Şekil 4. Çalışma alanına ait a. karar ağaçları ve b. kanonik korelasyon orman algoritmaları ile üretilen tematik haritalar.

5. Sonuçlar

Uydu görüntülerinin sınıflandırılması yardımıyla tematik harita üretim süreci birbirini takip eden bir dizi işlem adımından oluşmaktadır. Bu kapsamda uygun bir sınıflandırma algoritmasının kullanımı sınıflandırma işleminin başarısı ve dolayısıyla üretilen tematik haritanın doğruluğu açısından büyük önem arz etmektedir. Toplu öğrenme algoritmaları karmaşık yapıları sınıflandırma problemlerinin çözümündeki etkinliği ve sınıflandırma doğruluğu üzerinde pozitif etkileri nedeniyle uygulamalarda en çok tercih edilen kontrollü sınıflandırma algoritmaları arasındadır.

Bu çalışmada, toplu öğrenme algoritmalarının yeni bir üyesi olan kanonik korelasyon ormanı (KKO) algoritmasının multispektral Landsat OLI uydu görüntüsü üzerinden arazi kullanımı ve arazi örtüsü sınıflandırmasındaki performansı incelenmiştir. Algoritmanın sınıflandırma performansının karşılaştırılması amacıyla geleneksel en çok benzerlik (EÇB) ve popüler karar ağacı (KA) algoritmaları da değerlendirmeye alınmış ve sonuçlar analiz edilmiştir. Çalışma sonuçları, tematik harita üretiminde KKO algoritmasının istatistiksel EÇB ve popüler KA algoritmalarına göre daha başarılı sonuçlar verdiğini göstermektedir. Nitekim, sınıflandırma işleminde KKO algoritması kullanımıyla tematik harita doğruluğunda %11'e varan seviyelerde artış görülmüştür. Wilcoxon işaretli sıralar testi sonuçlarına göre sınıflandırma doğrulukları arasındaki bu düzeydeki farkın, %95 güven aralığında istatistiksel olarak anlamlı olduğu belirlenmiştir. Diğer bir ifadeyle, KKO algoritması kullanımıyla EÇB ve KA algoritmalarına göre sınıflandırma doğruluğunda istatistiksel olarak anlamlı artış sağlanmıştır. Çalışma sonuçları KA algoritmalarının tekil kullanımı yerine karar ağaçlarını esas alan karar ormanı tipli toplu öğrenme algoritmalarının kullanımının sınıflandırma doğruluğu üzerindeki pozitif ve istatistiksel olarak anlamlı etkisini ortaya koyarak, literatürdeki bulguları doğrular niteliktedir. Hesaplanan kullanıcı ve üretici doğrulukları

açısından sonuçlar ele alındığında, KKO algoritmasının benzer spektral özelliklere sahip ürün çeşitlerinin veya arazi örtüsü sınıflarının ayırt edilmesi noktasında daha başarılı olduğu görülmüştür. Üretilen tematik haritaların görsel olarak analizi de elde edilen bu sonucu doğrular niteliktedir. Tüm sonuçlar bir arada değerlendirildiğinde, yeni nesil bir karar ormanı olarak önerilen KKO algoritmasının, uzaktan algılanmış görüntülerin sınıflandırılmasında klasik sınıflandırma algoritmalarına göre etkili ve önemli bir alternatif olduğu görülmüştür.

6. Kaynaklar

- Belgiu, M. and Drăgut, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, **114**, 24-31.
- Briem, G.J., Benediktsson, J.A. and Sveinsson, J.R., 2002. Multiple classifiers applied to multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, **40(10)**, 2291-2299.
- Çölkesen, İ., 2017. Fonksiyonel ağaçlar ile obje-tabanlı sınıflandırma: WorldView-2 uydu görüntüsü örneği, *Harita Dergisi*, **157(1)**, 9-21.
- Colkesen, I. and Kavzoglu, T., 2017. The use of logistic model tree (LMT) for pixel- and object-based classifications using high-resolution WorldView-2 imagery. *Geocarto International*, **32(1)**, 71-86.
- Çölkesen, İ. ve Kavzoğlu, T., 2010. Farklı boyutta eğitim örnekleri için destek vektör makinelerinin sınıflandırma performansının analizi. 3. *Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu*. 11-13 Ekim, Gebze, Kocaeli.
- Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, **40(2)**, 139-157.
- Diez-Pastor, J.F., Rodriguez, J.J., Garcia-Osorio, C.I. and Kuncheva, L.I., 2015. Diversity techniques improve the performance of the best imbalance learning ensembles. *Information Sciences*, **325**, 98-117.
- Foody, G. M., Mathur, A., Sanchez-Hernandez, C., and Boyd, D.S., 2006. Training set size requirements for the classification of a specific class. *Remote Sensing of Environment*, **104(1)**, 1-14.
- Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forests for land cover classification. *Pattern Recognition Letters*, **27(4)**, 294-300.
- Ghosh, A. and Joshi, P.K., 2014. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, **26**, 298-311.
- Huang, C., Chen, Y., Zhang, S., Li, L., Shi, K. and Liu, R., 2016. Surface water mapping from Suomi NPP-VIIRS imagery at 30 m Resolution via blending with Landsat data. *Remote Sensing*, **8(8)**, 631.
- Jurek, A., Bi, Y.X., Wu, S.L. and Nugent, C., 2014. A survey of commonly used ensemble-based classification techniques. *Knowledge Engineering Review*, **29(5)**, 551-581.
- Kavzoğlu, T. ve Çölkesen, İ., 2010. Karar ağaçları ile uydu görüntülerinin sınıflandırılması: Kocaeli örneği, *Harita Teknolojileri Elektronik Dergisi*, **2**, 36-45.
- Kavzoglu, T. and Colkesen, I., 2012. The effects of training set size for the performance of support vector machines and decision trees. *Proceedings of 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 10-13 July, Florianópolis, SC, Brazil.
- Kavzoglu, T. and Colkesen, I., 2013. An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping. *International Journal of Remote Sensing*, **34(12)**, 4224-4241.
- Kavzoglu, T. and Colkesen, I., Yomralioglu, T., 2015. Object-based classification with rotation forest ensemble learning algorithm using very-high-resolution WorldView-2 image. *Remote Sensing Letters*, **6(11)**, 834-843.
- Kavzoglu, T. and Mather, P.M., 2003. The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, **24(23)**, 4907-4938.

- Li, M., Zang, S.Y., Zhang, B., Li, S.S. and Wu, C.S., 2014. A review of remote sensing image classification techniques: The role of spatio-contextual information. *European Journal of Remote Sensing*, **47**, 389-411.
- Lu, D. and Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, **28(5)**, 823-870.
- Mather, P.M., 1999. *Computer Processing of Remotely Sensed Images*, Second Edition, Chichester: John Wiley & Sons.
- Opitz, D.W. and Maclin, R.F., 1997. An empirical evaluation of bagging and boosting for artificial neural networks. In *International Conference on Neural Networks*, (1401-1405 vol.1403).
- Pal M. and Mather P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, **86(4)**, 554-565.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, **26(1)**, 217-222.
- Quinlan J.R., 1993. *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA.
- Rainforth, T. and Wood F., 2015. Canonical correlation forests. <https://arxiv.org/pdf/1507.05444.pdf>. Erişim 22.03.2017.
- Ren, Y., Zhang, L. and Suganthan, P.N., 2016. Ensemble Classification and Regression-Recent Developments, Applications and Future Directions. *IEEE Computational Intelligence Magazine*, **11(1)**, 41-53.
- Rokach, L., 2010. Ensemble-based classifiers. *Artificial Intelligence Review*, **33(1-2)**, 1-39.
- Safavian S.R. and Landgrebe D., 1991. A survey of decision tree classifier methodology, *IEEE Transactions on Systems Man and Cybernetics*, **21(3)**, 660-674.
- Tso B., and Mather, P.M., 2009. *Classification Methods for Remotely Sensed Data*. Second edition, Boca Raton, FL: CRC Press.
- Xia, J., Yokoya, N. and Iwasaki, A., 2017. Hyperspectral image classification with canonical correlation forests. *IEEE Transactions on Geoscience and Remote Sensing*, **55(1)**, 421-431.
- Zhang, H., and Roy, D., 2016. Computationally Inexpensive Landsat 8 Operational Land Imager (OLI) Pansharpening, *Remote Sensing*, **8(3)**, 180.