

KANONİK KORELASYON ORMAN ALGORİTMASI İLE UZAKTAN ALGILANMI GÖRÜNTÜLERİN SINIFLANDIRILMASI

. Çölkesen *, T. Kavzoğlu

Gebze Teknik Üniversitesi, Harita Müh. Bölümü, 41400 Gebze-Kocaeli
(icolkesen@gtu.edu.tr, kavzoglu@gtu.edu.tr)

ANAHTAR SÖZCÜKLER: Sınıflandırma, Toplu Öğrenme, Kanonik Korelasyon Orman, Karar Ağaçları, En Çok Benzerlik

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 karmaık ve gürültülü veri setleri kullanımındaki başarıları nedeniyle birçok sınıflandırma alanında yaygın olarak kullanılmaktadır. Kanonik korelasyon ormanı (KKO) algoritması son yıllarda sınıflandırıcıların performansının artırılması amacıyla yeni nesil bir karar ormanı olarak önerilen bir toplu öğrenme algoritmasıdır. Bu çalışmada, KKO algoritmasının arazi örtüsü ve arazi kullanımının haritalandırılmasındaki sınıflandırma performansının araştırılmasında temel veri kaynağı olarak Temmuz 2016 tarihinde elde edilen multispektral Landsat OLI görüntüsü kullanılmıştır. Yöntemin sınıflandırma performansı popüler en çok benzerlik (EÇB) ve karar ağaçları (KA) algoritmaları ile karşılaştırılmış ve istatistiksel olarak analiz edilmiştir. Elde edilen sonuçlar, genel sınıflandırma doğruluklarına göre KKO algoritmasının EÇB ve KA algoritmalarından daha üstün olduğunu göstermektedir. Sınıflandırma performansları karşılaştırıldığında KKO algoritması ile %11'e varan yüksek sınıflandırma doğrulukları elde edildiği ve sınıflandırma performansları arasındaki farklılıkların istatistiksel olarak anlamlı olduğu görülmüştür. Sonuç olarak, bu çalışma KKO algoritmasının multispektral uydu görüntülerinin sınıflandırılmasındaki etkinliğini ortaya koymaktadır.

CLASSIFICATION OF REMOTELY SENSED IMAGERY WITH CANONICAL CORRELATION FOREST ALGORITHM

KEYWORDS: Classification, Ensemble Learning, Canonical Correlation Forest, Decision Trees, Maximum Likelihood

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 as 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 forests (CCFs) has been recently introduced as a new type of decision forest to improve the individual classifier performances. In this study, multispectral Landsat OLI image acquired in July 2016 was used as a main data source to investigate the classification performance of CCFs algorithm in land use and land cover mapping. The classifier performances were compared and statistically analyzed with popular maximum likelihood (ML) and decision tree (DT) classifiers. Results showed that CCFs algorithm outperformed the ML and DT algorithms in terms of overall accuracy. When the classification performances were compared, it was observed that higher classification accuracies up to 11% obtained with the CCFs algorithm and the differences in the classification performances were found to be statistically significant. All in all, this study revealed the effectiveness of CCFs algorithm in the classification of multispectral satellite imagery.

1. GİRİŞ

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, deyim analizi, tarımsal içerikli araştırmalar ve arazi kullanımı planlamaları gibi birçok uygulamada temel altlık olarak kullanılmaktadır. Bu nedenle tematik haritaların doğru ve yeryüzü özelliklerinin temsiliyeti noktasındaki güvenilirliği söz konusu çalışmaları için büyük önem arz etmektedir. Sınıflandırma alanının başlıca tematik harita

doğruluğu ile doğrudan ilişkisi olup; ilgi duyulan alandaki yeryüzü özellikleri, uzaktan algılanan görüntülerin çözünürlüğü ve uygun bir sınıflandırma yönteminin seçimini gibi çeşitli faktörelere 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 karmaık sınıflandırma probleminin modellenmesi için istatistiksel kabullere dayanmayan sınıflandırma algoritmalarının kullanımı artı

* Tel: 0.262.605.1808

göstermi tir. Bu algoritmalar arasında yapay sinir a ları, destek vektör makineleri, karar a açları, hibrit sınıflandırma modelleri (örne in, fonksiyonel model a açlar ve lojistik model a açlar) ve toplu ö renme algoritmaları (örne in, rastgele orman ve rotasyon orman) sınıflandırma do rulu undaki i barıları nedeniyle literatürde yaygın olarak kullanılmaktadır (Kavzoglu and Mather, 2003; Belgiu and Dr gu , 2016; Colkesen and Kavzoglu, 2017; Çölkesen, 2017).

Sınıflandırma algoritmalarının birle tirilmesi veya kombinasyonu olarak da bilinen toplu ö renme algoritmalarının sınıflandırma ve regresyon problemlerinin çözümünde kullanımı makine ö renmesi veya veri madencili i alanında önemli bir ara tırma konusudur (Dietterich, 2000; Briem et al., 2002; Rokach, 2010; Jurek et al., 2014). Toplu ö renme algoritmalarının temel çalı ma prensibi, birden fazla sınıflandırıcının bir arada kullanımı ve her bir sınıflandırıcının tahminlerinin birle imi neticesinde sınıflandırma i leminin gerçekte tirilmesidir (Dietterich, 2000). Bu yakla ımla tek bir sınıflandırma algoritması kullanımına göre daha yüksek sınıflandırma do rululuklarının elde edildi i birçok çalı ma ifade edilmi tir (Opitz and Maclin, 1999; Gislason et al., 2006; Kavzoglu and Colkesen, 2013). Birden çok karar a acı sınıflandırıcısının bir arada kullanılması esasına dayanan ve karar ormanı olarak da isimlendirilen toplu ö renme algoritmaları farklı karakteristiklere sahip uydu görüntülerinin sınıflandırmasında ba arıyla kullanılmaktadır (Pal, 2005; Kavzoglu and Colkesen, 2013; Kavzoglu et al., 2015). Karar a açlarını esas alan toplu ö renme algoritmalarının çalı ma prensibi orman yapısını olu turan her bir karar a acının farklı bir veri seti kullanılarak e itilmesi ve sınıflandırma sonuçlarının birle imi neticesinde sonuç tahminin yapılmasıdır. Bu açıdan ele alındı ında toplu ö renme algoritmasının ba arısı farklı e itim verilerinin olu turulması veya karar ormanı içerisindeki çe itlili in sa lanmasıyla do rudan ili kilidir (Diez-Pastor et al., 2015; Ren et al., 2016). Karar ormanındaki çe itlili in artırılması ve dolayısıyla sınıflandırma do rulu nun artırılmasına yönelik olarak geli tirilen kanonik korelasyon ormanı (KKO) yeni nesil bir toplu ö renme algoritmasıdır. Bu toplu ö renme modelinde a aç yapıları arasındaki de i kenli in artırılması amacıyla, kanonik korelasyon analizi yardımıyla hesaplanan bile enler kullanılarak karar a açlarının olu turulması söz konusudur (Rainforth and Wood, 2015).

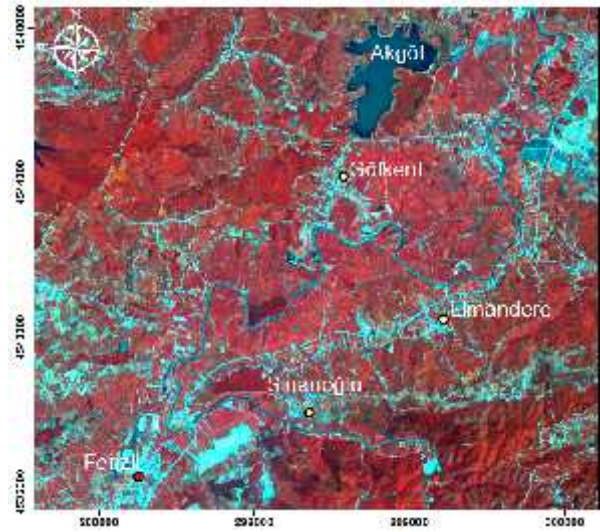
Bu çalı ma mada, KKO algoritmasıyla 2016 tarihinde kaydedilmi bir Landsat OLI görüntüsü kullanılarak, Sakarya ili Ferizli ilçesini içerisine alan ve tarımsal arazilerin ço unlukta oldu u çalı ma alanına ait arazi örtüsü ve arazi kullanımını temsil eden tematik harita üretilmesi amaçlanmı tir. Algoritmanın sınıflandırma performansı en çok benzerlik (EÇB) ve karar a açları (KA) sınıflandırma algoritmaları ile kar ıla tırılmı ve istatistiksel olarak analiz edilmi tir.

2. ÇALI MA ALANI VE KULLANILAN VER

Bu çalı ma mada Sakarya ilinin kuzeyinde yer alan Ferizli ilçe sınırları içerisindeki Sakarya nehrinin çevreledi i Sinano lu, Limandere ve Akgöl mahallelerini kapsayan bölge çalı ma alanı olarak belirlenmi tir (ekil 1). Çalı ma alanı ekonomik de eri olan ba ta mısır, fındık ve ekerpancarı olmak üzere birçok tarımsal ürün çe idinin yeti tildi i yakla ık 2.000 hektarlık bir alanı kapsamaktadır. Bölgede temel bitkisel üretim mısır ve fındık üzerine yo unlu mı tir. 2015 tarımsal üretim istatistiklerine göre Sakarya ili fındık üretiminde dördüncü, mısır üretiminde ise sekizinci sırada yer almaktadır (www.tuik.gov.tr). Bununla birlikte çalı ma sahası içerisinde

ekerpancarı, yonca ve çe itli bahçe bitkileri de (domates, karpuz, fasulye, biber vb.) yeti tirilmektedir.

Bu çalı ma mada, bölgeye arazi kullanımı ve arazi örtüsünün belirlenmesi amacıyla 2016 tarihinde kaydedilen Landsat OLI uydu görüntüsü kullanılmı tir. Uydu görüntüsünün geometrik düzeltilmesi ve UTM koordinat sistemine dönü ümünde mevcut hâlihazır haritalar ve bölgeye ait yüksek çözünürlüklü görüntülerden yararlanılmı tir. Mevcut uydu görüntüsünün 30m çözünürlü ündeki mavi, ye il, kırmızı, yakın kızıl ötesi ve yakın kızıl ötesi bantlarından olu an 6 multispektral bantlı 15m çözünürlü ündeki pankromatik bantla kayna tırılarak 15m çözünürlü ünde multispektral bantlar elde edilmi tir. Kayna tırma i leminin gerçekte tirilmesinde Gram-Schmidt algoritması kullanılmı , 3x3 görüntü yumu atma filtresi uygulanarak, yeniden örnekleme i leminde en yakın kom uluk enterpolasyon tekni inden yararlanılmı tir.



ekil 1. Çalı ma alanı konumu ve kullanılan uydu görüntüsü.

3. SINIFLANDIRMA ALGOR TMALARI

Sınıflandırma i lemi görüntü üzerindeki her bir piksele ait yansıma ve parlaklık de erleri kullanılarak piksellerin kullanıcı tarafından belirlenen sınıflara atanması i lemidir. Sınıflandırma sonucunda farklı arazi örtüsü ve kullanımlarını temsil eden tematik haritalar üretilmekte ve elde edilen haritalar yeryüzü özelliklerinin yorumlanması ve analizinde kullanılmaktadır. Bu çalı ma mada uydu görüntüsünün sınıflandırılması ve tematik harita üretiminde kanonik korelasyon ormanı, karar a açları ve en çok benzerlik sınıflandırıcısı kullanılmı tir.

3.1. En Çok Benzerlik Algoritması

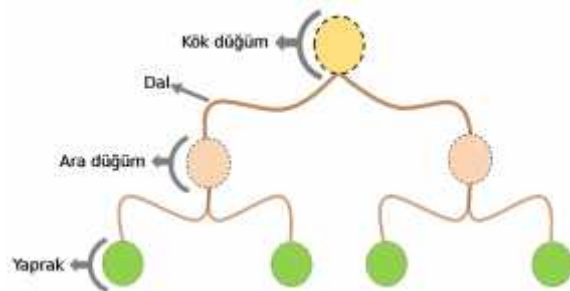
En çok benzerlik (EÇB) yöntemi uzaktan algılanmı görüntülerin sınıflandırılmasında yaygın olarak kullanılan istatistiksel bir sınıflandırma yöntemidir. Yöntem, sınıflandırma i leminin gerçekte tirilmesinde ortalama de er, varyans ve kovaryans gibi istatistik de erleri dikkate almaktadır. EÇB sınıflandırıcısı, e itim örneklerini olu turan sınıflar için olasılık fonksiyonları hesaplar ve her bir pikselin hangi sınıfa daha yakın oldu unu tespit eder. Bir piksele ait sınıf etiketi hesaplanan olasılıklara göre, pikselin en yüksek olasılıklı sınıfa atanması ile belirlenir. EÇB yönteminde, e itim verilerinin ve sınıfları olu turan piksellerin normal da ılımda oldu u kabul edilir. Sınıflandırma i leminin gerçekte tirilmesinde, her sınıf

için olu turacak varyans-kovaryans matris de erlerini ve piksellerin özellik uzayındaki da ılımlarını dikkate alınmaktadır (Mather, 1999).

3.2. Karar Aaçları Algoritması

Karar aaçları (KA) matematiksel olarak kolay ve anlaşılabilir yapıyla son yıllarda arazi kullanımı ve sınıflandırma problemlerinde sıklıkla kullanılan bir yöntemdir (Kavzo lu ve Çölkesen, 2010). KA arazi örtüsünün sınıflandırılması i leminde çok amaçlı veya ardı k bir yaklaşım tekni i kullanılmaktadır. Yöntemin temel i lem yapısı genel olarak karma k yapıdaki bir sınıflandırma probleminin çözülmesinde çok amaçlı bir hale getirerek basit bir karar verme i lemi gerçekleştirir (Safavian and Landgrebe, 1991; Kavzo lu ve Çölkesen, 2010).

Karar aacının temel yapısı düüm, dal ve yaprak olarak adlandırılan üç temel kısımdan oluşur (ekil 2). Bu aaç yapısında her bir öznelik bir düüm tarafından temsil edilir. Dallar ve yapraklar aaç yapısının di erelemanlarıdır. Aaçta en son kısım yaprak en üst kısım ise kök olarak adlandırılır. Kök ve yapraklar arasında kalan kısımlar ise dal olarak ifade edilir. E itim verileri ve bu verilere ait öznelik bilgilerinden yararlanılarak bir karar aacı yapısı olu umunda temel prensip verilere ili kin bir dizi sorular sorulması ve elde edilen cevaplar do rultusunda karar kuralları olu turulmasıdır. Aacın ilk düümü olan kök düümünde verilerin sınıflandırılması ve aaç yapısının olu turulması için sorular sorulmaya ba lanır ve dalları olmayan düümler ya da yapraklara gelene kadar bu i lem devam eder (Quinlan, 1993; Pal and Mather, 2003).

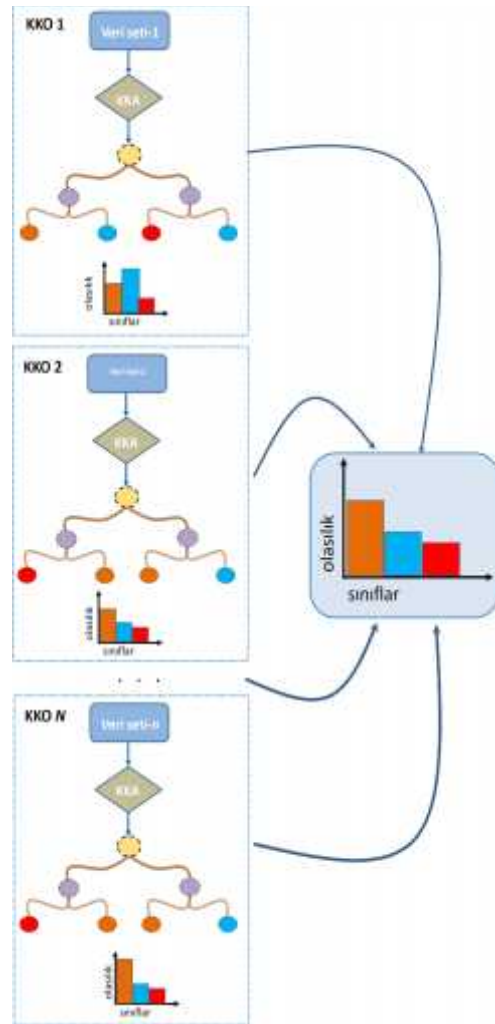


ekil 2. Basit bir karar aacı algoritması yapısı.

3.3. Kanonik Korelasyon Orman Algoritması

Kanonik korelasyon ormanı (KKO) karar aaçlarını esas alan ve kontrollü sınıflandırma i leminin gerçekleştirilmesi için tasarlanan yeni nesil bir toplu öğrenme algoritmasıdır (Rainforth and Wood, 2015, Xia et al., 2017). Karar aaçlarını 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 ekil 3'den de görülece i üzere bir dizi kanonik korelasyon aacının bir araya geldi i bir karar ormanı olu turmaktır. Bu karar ormanı modelindeki her bir aaç kanonik korelasyon (KK) analizi sonucunda elde edilen bile enler kullanılarak e itilir. KK analizi iki veya daha fazla di ken arasındaki ili kilerin incelenmesinde kullanılan bir dönüüm tekni i olup, KKO algoritması olu umunda e itim veri setindeki sınıflar ile özellikler (bantlar) arasındaki korelasyonu maksimum olmasını sa layan özellik dönüümleri gerçekleştirilir. Dönüüm sonrasında elde edilen bile enler yardımıyla aacın dallanması ve geli mi i lemleri gerçekleştirilir. Klasik karar aacı modeli sınıflar arasında eksen

bazlı (axis aligned) bölümlenmeler ile karar sınırları tanımlayarak sınıflandırma i lemi gerçekleştirilen, kanonik korelasyon aacı KK dönüümü ile olu an özellik uzayında sınıflar arasında e imli (oblique) karar sınırları tanımlayabilmektedir. Sahip olduğu bu özelliklerle KKO algoritması rastgele orman ve rotasyon orman algoritmalarının geli mi bir versiyonu olarak ifade edilmektedir. Toplu öğrenme modelindeki aaç sayısı karar ormanı algoritmalarının tamamı için kullanıcı tarafından belirlenen bir parametre olmasına rağmen, söz konusu parametredeki di imler KKO algoritmasının sınıflandırma performansında istatistiksel olarak anlamlı di imler ortaya çıkarmadı ı ifade edilmektedir (Rainforth and Wood, 2015). Dolayısıyla, rastgele orman algoritmasının aksine, KKO algoritmasının çok daha az sayıda aaç ile sonuca ula şabilmektedir. Ormandaki her bir kanonik korelasyon aacı sınıflandırma sonucu bir oy alır ve orman içerisindeki tüm aaçlardan en çok oy alan (en düşük hata oranına sahip) belirlenerek sınıflandırmaya esas aaç yapısı tespit edilir. Sınıf etiketi bilinmeyen herhangi bir piksel, tüm aaç tahminlerinde en fazla oy alan sınıfa atanması suretiyle sınıflandırılır.



ekil 3. Kanonik korelasyon ormanı çalışma prensibi.

4. UYGULAMA

Çalışma alanında yapılan detaylı arazi çalışmaları sonucunda bölgede 9 temel arazi örtüsü/kullanım sınıfı olduğuna karar verilmiştir. Arazi örtüsü olarak orman, bozkır, su ve toprak sınıfları tanımlanmıştır. Bölge tarımsal ürün çe itlili i 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 (eker pancarı, yonca vb.) tarım sınıfı olarak adlandırılmış ve tek bir sınıf altında değerlendirilmeye alınmıştır. Bölge içerisindeki asfalt yollar ve yapılaşma alanları için ayrı ayrı yol ve bina sınıfları tanımlanmıştır. Belirlenen örnekleme alanları uydu görüntüleri ile birlikte tirilerek sınıflandırma işleminde kullanılmak üzere rastgele örnekleme prensibi esas alınarak eğitim ve test veri setleri oluşturulmuştur. Hesaplanan genel sınıflandırma doğruluklarının karşılaştırılmasında objektiflik sağlamak amacıyla eğitim ve test veri setleri belirlenirken eğitim sayısında örnek seçilmiştir. Sonuç olarak belirlenen 9 arazi örtüsü sınıfı için eğitim verisi olarak 4.050 piksel (her bir sınıf için 450 piksel), 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 üretimi için 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 işleminde tek denetimli karar ağaç algoritmasından C4.5 algoritması kullanılmıştır. Açık model oluğunda dallanmaya esas olacak bantların seçiminde kazanç oranı kriteri, oluşturulan karar ağacının sadeleştirilmesi ve sınıflandırma doğruluğunun artırılması amacıyla ön budama yöntemi kullanılmıştır. Eğitim veri seti kullanılarak oluşturulan karar ağaç modelinde 366 düğüm ve sınıf etiketlerini içeren 367 yaprak bulunmaktadır. Açık modelin kök düğümünde dallanmaya esas olacak özellik olarak Landsat OLI'nin ikinci orta kızılötesi (SWIR-2) bandının seçildiği, kök düğümünden 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 oluğunda 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 çalışmalar söz konusu parametrede denetimlerin sınıflandırma doğruluğu üzerinde istatistiksel olarak anlamlı denetimler ortaya çıkarmadığı, çok az sayıda kanonik korelasyon ağacında oluşturulan 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. Oluşturulan KKO modelinin geçerliliği test veri seti kullanılarak analiz edildiğinde, Tablo 1'den de görüleceği üzere hesaplanan genel doğruluk ve Kappa değeri %91,35 ve 0,90'dır.

Uygulama kapsamında multispektral Landsat OLI görüntüsünün sınıflandırılması ve tematik harita üretiminde kullanımı değerlendirilmeye alınan KKO algoritmasının sınıflandırma performansı geleneksel EÇB ve KA algoritmaları ile elde edilen sonuçlarla karşılaştırıldı. 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 ağaç algoritması kullanımı yerine birden fazla ağaçtan oluşan toplu öğrenme algoritması kullanımının sınıflandırma doğruluğunda önemli derecelerde (yaklaşık %11) artışı ortaya çıkardığını göstererek, literatürde daha önce gerçekleştirilen çalışmalarda bu konu hakkındaki bulguları destekler niteliktedir.

Genel sınıflandırma doğruluklarının karşılaştırılmasının yanında oluşturulan hata matrisleri yardımıyla arazi örtüsü ve kullanım sınıflarına ilişkin kullanıcı ve üretici doğrulukları da hesaplanmıştır 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 doğrulukları %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 doğrulukları %9-25 arasında artışı 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.

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

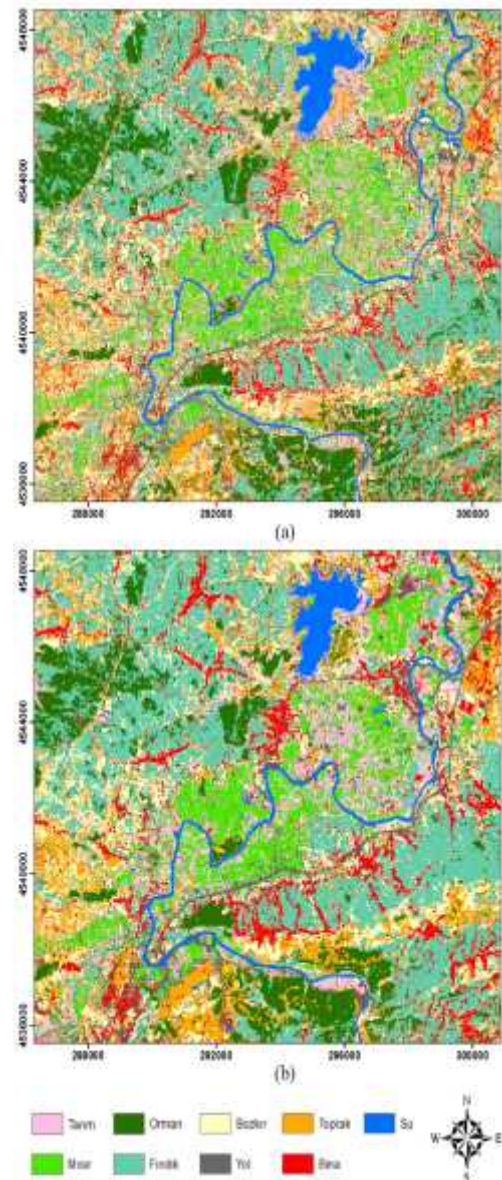
Değerlendirilmeye alınan algoritmaların sınıflandırma performansları arasındaki farklılıkların istatistiksel olarak

anamlılı mın tespiti amacıyla Wilcoxon i aretli sıralar testinden yararlanılmıştır. ki sınıflandırma algoritmasının performans farklılıklarının sıralanması esasına dayanan istatistik test, hesaplanan farkların pozitif ve negatif farklar olarak ayrı ayrı sıralanmasını ve daha sonra birbirleriyle karşılaştırılması esas almaktadır. Hesaplanan test istatisti i , belirlenen güven aralığında kritik normal dağılım tablosundan bulunmuş ve erlendirilmeye alınan iki sınıflandırma sonucu arasında istatistik olarak anlamlı bir fark olduğu sonucuna varılır (Çölkesen, 2017). 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 diğer iki yöntemle göre sınıflandırma doğruluğundaki %11’lik bir iyileşme istatistiksel olarak anlamlı bir artış olduğunu göstermektedir. Diğer taraftan, Wilcoxon test sonuçları EÇB ve KA algoritmalarının uygulama kapsamında kullanılan veri seti için istatistiksel olarak benzer sınıflandırma performansı sergilediklerini doğrulamaktadır. Aynı şekilde yeni bir toplu öğrenme algoritması olan KKO algoritmasının, tek bir karar ağacı kullanımına göre sınıflandırma sonuçlarında istatistiksel olarak anlamlı bir artış ortaya çıkardığı görülmektedir.

Tablo 3. Wilcoxon i aretli sıralar test sonuçları.

Karşılaştırma	Test istatisti i
EÇB - KA	0,71
KKO - EÇB	18,65
KKO - KA	19,36

Uygulamada de erlendirilmeye alınan tüm sınıflandırma algoritmalarına ait sınıflandırma modelleri uydu görüntüsüne uygulanarak çalışılma alanına ilişkin tematik haritalar üretilmiştir. Söz konusu tematik haritalara örnek olarak KA ve KKO algoritmalarıyla üretilen tematik haritalar ekil 4’de gösterilmiştir. KA ile üretilen tematik haritada (ekil 4a) literatürde tuz-biber (salt-pepper) olarak bilinen gürültü oranının daha fazla olduğu ve dolayısıyla sınıflara ait doğruların birbirinden ayırt edilmesinin güç olduğu görülmektedir. Özellikle çalışılma bölgesi içerisindeki canlı vejetasyonu gösteren mısır, fındık, orman ve diğer tarım alanlarını temsil eden pikseller arasındaki hatalı sınıflandırmalar neticesinde görüntü üzerinden ayırım yapmanın güçlüğü ifade edilebilir. Diğer taraftan KKO algoritması ile üretilen haritada (ekil 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 doğrular bir sonucu olarak ifade edilebilir. Su geçirimsiz yüzeyler olarak bilinen kentsel alanlar ve asfalt yollarına ait piksellere ilişkin sınıflandırma sonuçları incelendiğinde, KA ile üretilen tematik haritada (ekil 4a) söz konusu piksellerin birbirine karışması ve görsel analizinin güç olduğu görülmektedir. Buna karşın KKO ile üretilen tematik haritada (ekil 4b) asfalt ve beton yollarına ait piksellerin sınıflandırılmasının daha doğru olduğu ifade edilebilir.



ekil 4. Çalışılma alanına ait a) karar ağacı ve b) kanonik korelasyon ormanı 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 de erlendirilmeye

alınımı 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 ileminde KKO algoritması kullanımıyla tematik harita do rulu unda %11'e varan seviyelerde artış görülmüştür. Wilcoxon i aretli sıralar testi sonuçlarına göre sınıflandırma do rulukları arasındaki bu düzeydeki farkın, %95 güven aralığında istatistiksel olarak anlamlı olduğu belirlenmiştir. Diğer bir ifadeyle, KKO algoritması kullanımıyla EÇB ve KA algoritmalarına göre sınıflandırma do rulu unda istatistiksel olarak anlamlı artışa lanmıştır. Çalışma sonuçları KA algoritmalarının tekil kullanımı yerine karar ağaçlarını esas alan karar ormanı tipli toplu öğrenme algoritmalarının kullanımının sınıflandırma do rulu u üzerindeki pozitif ve istatistiksel olarak anlamlı etkisini ortaya koyarak, literatürdeki bulguları doğrular niteliktedir. Hesaplanan kullanıcı ve üretici do rulukları 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 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.

KAYNAKLAR

- Belgiu, M., Dragut, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp. 24-31.
- Briem, G.J., Benediktsson, J.A., Sveinsson, J.R., 2002. Multiple classifiers applied to multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(10), pp. 2291-2299.
- Çölkesen, İ., 2017. Fonksiyonel ağaçlar ile obje-tabanlı sınıflandırma: WorldView-2 uydu görüntüsü örneği, *Harita Dergisi*, 157(1), pp. 9-21.
- Colkesen, İ., 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), pp. 71-86.
- Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2), pp. 139-157.
- Diez-Pastor, J.F., Rodriguez, J.J., Garcia-Osorio, C.I., Kuncheva, L.I., 2015. Diversity techniques improve the performance of the best imbalance learning ensembles. *Information Sciences*, 325, pp. 98-117.
- Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forests for land cover classification. *Pattern Recognition Letters*, 27(4), pp. 294-300.
- Jurek, A., Bi, Y.X., Wu, S.L., Nugent, C., 2014. A survey of commonly used ensemble-based classification techniques. *Knowledge Engineering Review*, 29(5), pp. 551-581.
- Kavzoglu, T., Çölkesen, İ., 2010. Karar ağaçları ile uydu görüntülerinin sınıflandırılması: Kocaeli örneği, *Harita Teknolojileri Elektronik Dergisi*, 2, pp. 36-45.
- Kavzoglu, T., Colkesen, İ., 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), pp. 4224-4241.
- Kavzoglu, T., Colkesen, İ., Yomralioglu, T., 2015. Object-based classification with rotation forest ensemble learning algorithm using very-high-resolution WorldView-2 image. *Remote Sensing Letters*, 6(11), pp. 834-843.
- Kavzoglu, T., Mather, P.M., 2003. The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, 24(23), pp. 4907-4938.
- Li, M., Zang, S.Y., Zhang, B., Li, S.S., Wu, C.S., 2014. A review of remote sensing image classification techniques: The role of spatio-contextual information. *European Journal of Remote Sensing*, 47, pp. 389-411.
- Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5), pp. 823-870.
- Mather, P.M., 1999. *Computer Processing of Remotely Sensed Images*, Second Edition, Chichester: John Wiley & Sons.
- Opitz, D.W., Maclin, R.F., 1997. An empirical evaluation of bagging and boosting for artificial neural networks. In *International Conference on Neural Networks*, (pp. 1401-1405 vol.1403).
- Pal M., Mather P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4), pp. 554-565.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), pp. 217-222.
- Quinlan J.R., 1993. *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA, 302 p.
- Rainforth, T., Wood F., 2015. Canonical correlation forests. <https://arxiv.org/pdf/1507.05444.pdf>. Erişim 22.03.2017.
- Ren, Y., Zhang, L., Suganthan, P.N., 2016. Ensemble Classification and Regression-Recent Developments, Applications and Future Directions. *IEEE Computational Intelligence Magazine*, 11(1), pp. 41-53.
- Rokach, L., 2010. Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2), pp. 1-39.
- Safavian S.R., Landgrebe D., 1991. A survey of decision tree classifier methodology, *IEEE Transactions on Systems Man and Cybernetics*, 21(3), pp. 660-674.
- Tso B., Mather, P.M., 2009. *Classification Methods for Remotely Sensed Data*. Second edition, Boca Raton, FL: CRC Press.
- Xia, J., Yokoya, N., Iwasaki, A., 2017. Hyperspectral image classification with canonical correlation forests. *IEEE Transactions on Geoscience and Remote Sensing*, 55(1), pp. 421-431.