HİPERSPEKTRAL GÖRÜNTÜLERİN SINIFLANDIRILMASINDA GELİŞMİŞ KARAR ORMANI TOPLU ÖĞRENME ALGORİTMALARININ PERFORMANS ANALİZİ

İsmail ÇÖLKESEN¹, Ömer Habib ERTEKİN²

¹ Dr. Öğr. Üyesi, Gebze Teknik Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, 41400, Gebze, KOCAELİ, icolkesen@gtu.edu.tr

²Müh., Gebze Teknik Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, 41400, Gebze, KOCAELİ, habibertekin@gtu.edu.tr

ÖZET

Günümüze kadar sınıflandırma işleminin gerçekleştirilmesi ve güvenilirliği yüksek tematik harita üretimi amacıyla birçok sınıflandırma algoritması geliştirilmiş ve uygulamalarda kullanılmıştır. Çoklu sınıflandırma sistemleri olarak bilinen ve sınıflandırma işleminin gerçekleştirilmesinde birden çok sınıflandırma algoritmasının bir arada kullanılmasını esas alan toplu öğrenme algoritmaları uydu görüntülerinin sınıflandırılmasını da içerisinde alan birçok görüntü tanıma probleminde başarıyla kullanılmaktadır. Çoklu sınıflandırma sistemleri içerisinde işlem hızı, tahmin başarısı ve kullanıcılar açısından kolay anlaşılabilir yapısı nedeniyle karar ağaçlarını esas alan toplu öğrenme algoritmaları veya karar ormanı olarak bilinen algoritmalar yaygın bir kullanıma sahiptir. Bu çalışmada, son yıllarda farklı disiplinlerce sınıflandırma problemlerinde kullanılmaya başlanan ancak uzaktan algılanmış görüntülerin sınıflandırılması problemlerindeki performansları sınırlı sayıda çalışmada test edilen yeni nesil karar ormanı yapılarından rotasyon orman (RotOr) ve kanonik korelasyon orman (KKO) algoritmalarının hiperspektral görüntülerin sınıflandırılmasındaki performansları karşılaştırılmalı olarak analiz edilmiştir. Söz konusu algoritmalar ile birlikte standart karar ormanı algoritması olarak bilinen rastgele orman (RO) algoritması da değerlendirmeye alınmış ve tüm algoritmaların performansları istatistiksel olarak analiz edilmiştir. Uygulamada temel veri kaynağı olarak AVIRIS sensöründen elde edilen 224 spektral banda sahip Indian Pine veri seti ve ROSIS sensöründen elde edilen 103 spektral banda sahip Pavia Üniversitesi hiperspektral veri setleri kullanılmıştır. Algoritmaların performans analizinde standart doğruluk ölçütlerinin (genel doğruluk ve Kappa değerleri) yanında, işlem süresi ve karar ormanı sınıflandırıcıları için önemli bir performans göstergesi olarak kabul edilen çeşitlilik ölçütleri (diversity measures) değerlendirmeye alınarak algoritmaların performansları detaylı bir şekilde analiz edilmiştir. Elde edilen sonuçlar her iki hiperspektral görüntü için de KKO algoritması ile en yüksek sınıflandırma doğruluğuna ulaşıldığını (AVIRIS için %87 ve ROSIS için %94) ve sınıflandırma doğrulukları arasındaki farklılıkların McNemar's testine göre istatistiksel olarak anlamlı olduğu görülmüştür. Sonuç olarak, bu çalışma RO algoritmasının gelişmiş versiyonları olarak bilinen KKO ve RotFor algoritmalarının hiperspektral görüntülerin sınıflandırılmasındaki etkinliğini ortaya koymaktadır.

Anahtar Sözcükler: hiperspektral görüntü, kanonik orman, rastgele orman, rotasyon orman, toplu öğrenme

ABSTRACT

PERFORMANCE ANALYSIS OF ADVANCED DECISION FOREST ENSEMBLE LEARNING ALGORITHMS IN HYPERSPECTRAL IMAGE CLASSIFICATION

Up to now, a variety of classification algorithms have been developed and utilized in applications for performing classification process and producing accurate thematic maps. Ensemble learning algorithms, known as multiple classifier systems and based on the use of multiple-classifiers to perform classification process, have been successfully used in many pattern recognition problems including classification of satellite imagery. Within the multiple-classifier systems, ensemble learning algorithms based on decision trees or known as decision forest algorithms have been widely-used due to their processing speed, predictive capacity and easy structure to understand by the user. In this study, the comparative performance of up-to-date decision forest algorithms, rotation forest (RotFor) and canonical correlation forest (CCF) which have been used for classification problems in different disciplines, but they have limited use in remote-sensing applications, were compared and analysed in the classification of hyperspectral imagery. In this study, the performances of up-to-date decision forest algorithms, rotation forest (RotFor) and canonical correlation forest (CCF) which have been used for classification problems in different disciplines, but they have limited use in remote-sensing applications, were compared and analysed in the classification of hyperspectral imagery. In application, the Indian Pine data set with 224 spectral bands obtained from the AVIRIS sensor and the hyperspectral data set of Pavia University with 103 spectral bands obtained from the ROSIS sensor were used as the basic data source. In this study, as well as the standard accuracy measures (overall accuracy and Kappa values), the performances of the algorithms were analysed in detail considering computational cost and diversity measures accepted as an important performance indicator for decision forest algorithms. Results showed that the highest classification accuracies (87% for AVIRIS and 94% for ROSIS) were achieved with the CCF algorithm for both hyperspectral images and differences in classification accuracy were statistically significant according to McNemar's test. All in all, this study verifies the effectiveness of CCF and RotFor algorithms, known as advanced versions of the RO algorithm, in the classification of hyperspectral images.

Keywords: hyperspectral image, canonical forest, random forest, rotation forest, ensemble learning

1. GİRİŞ

Yeryüzündeki doğal ve yapay nesneleri ifade eden arazi kullanımı ve arazi örtüsüne ilişkin mekânsal dağılımların doğru ve güvenilir bir şekilde tespit edilmesi küresel ve yerel ölçekli birçok çalışma için büyük önem arz etmektedir. Geçmişten günümüze uzaktan algılama teknolojileri yeryüzüne ilişkin bilgilerin elde edilmesinde önemli bir araç olarak kullanılmış ve uzaktan algılanmış görüntüler birçok çalışmada temel veri kaynağı olarak değerlendirmeye alınmıştır. Bu bağlamda uydu görüntülerinin sınıflandırılması yardımıyla arazi örtüsü ve arazi kullanımını temsil eden tematik haritaların üretilmesi uzaktan algılama alanında önemli çalışma konularından birisidir. Sınıflandırma işlemi neticesinde üretilen tematik haritaların yeryüzündeki doğal ve yapay özelliklerin temsiliyeti noktasındaki güvenilirliği, arazi kullanımı ve arazi örtüsü bilgisine ihtiyaç duyan birçok çalışmanın başarısı açısından kritik öneme sahiptir. Tematik harita doğruluğu sınıflandırma işleminin başarısı ile doğrudan ilişkili olup; temel olarak ilgi duyulan alandaki arazi kullanımı ve arazi örtüsü yapısına, kullanılan görüntülerin özelliklerine, çalışma ölçeğine ve sınıflandırma işleminin gerçekleştirilmesinde kullanılacak yöntemle bağlı olarak değişkenlik göstermektedir (Lu ve Weng 2007). Sınıflandırma işleminin gerçekleştirilmesi amacıyla günümüze kadar birçok sınıflandırma tekniği ve yöntemi geliştirilmiş ve uygulamalarda kullanılmıştır (Tso ve Mather 2009; Li vd. 2014). Uzaktan algılama teknolojilerinde yaşanan yenilikler beraberinde uydu görüntülerinin mevcut özelliklerinin özellikle çözünürlükler açısından gelişmesine imkân sağlamıştır. Bunun doğal bir sonucu olarak güncel ancak karmaşık yapıya sahip yerleşimlerin sınıflandırılması ve doğruluğu yüksek tematik harita üretiminde klasik sınıflandırma yöntemleri yerine makine öğrenme ve yapay zekâ algoritmalarının kullanılması önemli bir araştırma konusu hâline gelmiştir (Çölkesen 2017; Ghamisi vd. 2017; Maxwell vd. 2018).

Son yıllarda literatürde tek bir sınıflandırıcı yerine birden çok sınıflandırıcının belirli bir sınıflandırma probleminin çözümünde kullanılmasını esas alan toplu öğrenme algoritmaları uzaktan algılama alanını da içerisine alan birçok sınıflandırma probleminin çözümünde kullanıldığı görülmektedir (Rokach 2010, Ren vd. 2016; Sagi ve Rokach 2018). Toplu öğrenme algoritmalarının esası birden çok sınıflandırıcının problem çözümünde kullanılması ve her bir sınıflandırıcının tahminlerinin ağırlıklı oylama yaklaşımı ile birleştirilerek nihai sınıflandırma sonucunun belirlenmesi şeklinde ifade edilebilir (Dietterich 2000). Literatürde toplu öğrenme algoritmaları kullanımı ile sınıflandırma performansında önemli derecede artış olduğu vurgulanmaktadır (Rokach 2010; Kavzoglu ve Colkesen 2013; Belgiu ve Drăguț 2016; Rokach 2016). Günümüze kadar toplu öğrenme algoritması olarak başta hızlandırma, torbalama, rastgele orman (RO) ve dönüşümlü orman (RotOr) olmak üzere birçok algoritma geliştirilmiştir. Birden çok karar ağacı sınıflandırıcısının bir arada kullanılması esasına dayanan ve karar ormanı olarak da isimlendirilen toplu öğrenme algoritmaları tahmin doğruluğundaki başarıları, kısa işlem süresi ve matematiksel olarak anlaşılabilir yapılarıyla yüzlerce spektral banttan oluşan ve karmaşık yapıya sahip hiperspektral uydu görüntülerinin sınıflandırılmasında başarıyla kullanılmaktadır. Örneğin, Chan ve Paelinckx (2008) 126 spektral banda sahip Hymap görüntüsünün sınıflandırılması probleminde karar ağaçlarını esas alan AdaBoost ve RO algoritmaları ile yapay sinir ağları algoritmasının sınıflandırma performanslarını karşılaştırmıştır. Elde edilen sonuçlar AdaBoost ve rastgele orman algoritmalarının benzer sonuçlar verdiği ama her ikisinin de yapay sinir ağlarından daha doğru sonuçlar elde ettiğini göstermektedir. Xia vd. (2014) AVIRIS, ROSIS ve DAIS sensörlerinden elde edilen hiperspektral görüntülerin sınıflandırılması probleminde torbalama, AdaBoost, RO, RotOr olarak bilinen toplu öğrenme algoritmaları yanında destek vektör makineleri algoritmasını kullanmış ve sınıflandırma sonuçlarına göre RO algoritmasıyla en yüksek sınıflandırma doğruluklarına ulaşıldığı belirtilmiştir. Bu çalışmada AVIRIS ve ROSIS sensörlerinden elde edilen hiperspektral görüntülerin sınıflandırılmasında gelişmiş karar ormanı algoritmalarından RotOr, kanonik korelasyon ormanı (KKO) ve RO algoritmalarının performansları karşılaştırılmalı olarak analiz edilmiştir.

2. ÇALIŞMA ALANI VE KULLANILAN VERİ

Çalışma kapsamında karar ağaçlarını esas alan RotFor, KKO ve RO toplu öğrenme algoritmalarının hiperspektral görüntülerin sınıflandırılmasındaki performansı karşılaştırılmalı olarak incelenmiştir. Bu amaç doğrultusunda birçok örüntü tanıma ve sınıflandırma algoritmasının performans analizinde test verisi olarak kullanılan AVIRIS ve ROSIS sensörleri tarafından kaydedilen iki farklı hiperspektral veri seti değerlendirmeye alınmıştır. Indian Pine veri seti, AVIRIS sensörü tarafından 0.4–2.5 µm aralığında kaydedilen 224 spektral banttan oluşmakta ve toplam 145x145 pikselden oluşan bir görüntüdür (Şekil 1a). Su soğurma bantları (104-108, 150-163) çıkarıldığında geriye kalan 200 bantlı veri seti sınıflandırma probleminde temel veri kaynağı olarak kullanılmıştır. Hiperspektral görüntünün üçte ikisinde tarım arazilerinden yer alırken, geriye kalan kısımlarını orman veya diğer bitki türleri oluşturmaktadır. Indian Pine görüntüsüne ait 16 arazi örtüsü sınıfını gösteren bir yer doğrulama verisi mevcuttur ve sınıflandırma işleminin gerçekleştirilmesinde söz konusu veri seti kullanılmıştır (Şekil 1b).

Şekil 1. (a) AVIRIS Indian pine veri seti ve (b) 16 arazi örtüsü sınıfını gösteren bir yer doğrulama verisi.

Pavia Üniversitesi veri seti İtalya’da Pavia bölgesini içerisine alan bölgede ROSIS sensörü tarafından kaydedilen 610x610 pikselden oluşan bir alanı kapsamaktadır (Şekil 2a). Görüntünün geometrik çözünürlüğü 1,3 metre olup toplam 103 spektral banda sahip bir görüntüdür. Hiperspektral görüntüdeki arazi örtüsünün büyük bir çoğunluğu su geçirimsiz yüzeylerden oluşurken diğer arazi örtüsü tipleri ağaçlar, çimenlik alanlar ve çıplak arazilerden oluşmaktadır. Şekil 2b’de gösterildiği üzere Pavia Üniversitesi görüntüsüne ait 9 arazi örtüsü sınıfını gösteren bir yer doğrulama verisi mevcuttur ve sınıflandırma işleminin gerçekleştirilmesinde söz konusu veri setinden yararlanılmıştır.

Şekil 2. (a) ROSIS Pavia üniversitesi veri seti ve (b) 9 arazi örtüsü sınıfını gösteren bir yer doğrulama verisi.

3. KARAR ORMANI TOPLU ÖĞRENME ALGORİTMALARI

Toplu öğrenme algoritmalarının temel çalışma prensibi belirli bir sınıflandırma probleminin çözümünde birden çok sınıflandırıcının kullanılması ve söz konusu sınıflandırıcıların tahminlerinden yola çıkılarak nihai kararın verilmesi şeklinde ifade edilebilir. Bu noktada, birçok farklı sınıflandırma algoritması kullanılarak toplu öğrenme modeli oluşturulabileceği gibi (hibrit model), tek bir sınıflandırıcı esas alınarak toplu öğrenme modeli (homojen model) oluşturabilmek mümkündür. Literatürde karar ormanı algoritmaları olarak isimlendirilen, karar ağaçlarını esas alan homojen yapıli toplu öğrenme algoritmaları uzaktan algılamada sınıflandırma problemlerinin çözümünde sıklıkla tercih edilmektedir. Karar ormanı algoritmalarının temel çalışma prensibi orman yapısını oluşturan her bir karar ağacının farklı bir veri seti kullanılarak eğitilmesi ve sınıflandırma sonuçlarının birleşimi neticesinde sonuç tahminin yapılmasıdır. Bu çalışmada, son yıllarda kullanılmaya başlanan KKO ve RotOr algoritmalarının hiperspektral görüntülerin sınıflandırması probleminde kullanımları ele alınmıştır. Sınıflandırma performansları yaygın kullanıma sahip RO algoritmasının performansı ile karşılaştırılarak sonuçlar analiz edilmiştir.

3.1. Rastgele Orman

Rastgele orman (RO) algoritması tahmin işlemindeki başarısından dolayı uzaktan algılanmış görüntülerin sınıflandırılmasında en yaygın kullanıma sahip toplu öğrenme algoritmalarındandır (Belgiu ve Drăguț 2016). RO algoritması, karar ağaçlarının temel sınıflandırıcı olarak kullanıldığı ve birden çok karar ağacının bir araya getirilerek toplu öğrenme modelinin oluşturulması esasına dayandığından karar ormanı olarak adlandırılan homojen yapıli bir toplu öğrenme yaklaşımdır (Breiman 2001). Şekil 3'te temel çalışma prensibi gösterilen RO algoritması, karar ormanındaki her bir ağaç modeli oluşturulmasında kullanılacak eğitim veri setini orijinal eğitim veri seti içerisinde rastgele seçilen veri setleri yardımıyla gerçekleştirir. Söz konusu rastgele örneklenen veri setinin 2/3'ü karar ağacı yapısını oluşturmak için kullanılırken 1/3'lük kısmı oluşturulan karar ağacı modelinin geçerliliğini test etmek için kullanılır. Ormandaki her bir ağaç yapısı rastgele seçilen veri setleri ile oluşturulduğundan belirli bir örnek için farklı tahminler üretebilmektedir. Bu durum ormandaki çeşitlilik (diversity) kavramıyla açıklanmaktadır. Bu nedenle, sınıf etiketi belli olmayan bir örneğin sınıflandırılması işleminde, karar ormanındaki her bir ağaç modelinin tahminleri ağırlıklı oylama (majority voting) prensibinden hareketle birleştirilerek örneğe ilişkin nihai sınıflandırma işlemi gerçekleştirilir. Diğer bir ifadeyle, sınıf etiketi bilinmeyen herhangi bir örnek, tüm ağaç tahminlerinde en fazla oy alan sınıfa atanması suretiyle sınıflandırılır.

Şekil 3. Rastgele orman algoritmasının temel çalışma prensibi.

3.2. Rotasyon orman

Rotasyon orman (RotOr) algoritması gelişmiş bir karar ormanı yaklaşımı olarak son yıllarda uzaktan algılanmış görüntülerin sınıflandırılmasında başarıyla kullanılmaktadır (Rodríguez vd. 2006; Kavzoglu ve Colkesen 2013; Zhang vd. 2018). RotOr algoritması temel sınıflandırıcı olarak karar ağaçlarını esas almakta ve ormandaki ağaç yapılarının oluşturulmasında rastgele seçilen veri setlerinin kullanılmasını esas aldığından RO algoritmasının gelişmiş bir versiyonu olarak ifade edilebilir. Aradaki en önemli fark ormandaki çeşitliliğin artırılmasına yönelik olarak, ağaç oluşumunda rastgele seçilen örneklerin direkt olarak kullanılması yerine, seçilen örneklerle ilk önce temel bileşenler analizinin uygulanması ve dönüşüm neticesinde elde edilen bileşenler yardımıyla ağaç yapılarının oluşturulmasıdır (Şekil 4). Diğer bir ifadeyle, ormandaki her bir karar ağacının eğitiminde kullanılacak veri seti temel bileşenler analizi yardımıyla belirlenmektedir. Temel bileşenler analizi ile ayırt ediciliği en yüksek olan özellikler (bantlar) belirlenmektedir. Veri setindeki değişkenliği arttırmak için dönüşüm neticesinde elde edilen tüm bileşenler dikkate alınır. Sınıf etiketi belli olmayan herhangi bir örneğe ilişkin nihai sınıflandırma işlemi ormandaki tüm ağaçlara ilişkin tahminlerin ağırlıklı oylama yaklaşımı ile birleştirilmesi neticesinde gerçekleştirilir.

Şekil 4. Rotasyon ormanı algoritmasının temel çalışma prensibi.

3.3. Kanonik Korelasyon Ormanı

Rainforth ve Wood (2015) tarafından önerilen kanonik korelasyon ormanı (KKO) karar ağaçlarını esas alan yeni nesil bir karar ormanı algoritmasıdır. Algoritmanın temel çalışma prensibi klasik karar ormanı yaklaşımlarına benzer şekilde olup, orman içerisindeki her bir karar ağacı modeli kanonik korelasyon analizi (KK) neticesinde elde edilen bileşenler kullanılarak oluşturulur (Şekil 5). Diğer bir ifadeyle, rastgele orman yönteminde olduğu gibi her bir ağaç modeli oluşumunda kullanılmak üzere orijinal veri seti içerisinde rastgele seçilen örneklerin kullanılması yerine, rastgele seçilen örneklere öncelikle KK analizi uygulanır ve elde edilen bileşenler ağaç oluşumunda dikkate alınır. Karar ağaçlarının eğitiminde ve model oluşumunda KK analizi ile elde edilen bileşenler kullanıldığından, oluşturulan ağaç modeli kanonik korelasyon ağacı olarak isimlendirilmektedir. 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 (Çölkesen ve Kavzoğlu, 2017).

Şekil 5. Kanonik korelasyon ormanı algoritmasının temel çalışma prensibi.

4. UYGULAMA

Uygulama kapsamında AVIRIS ve ROSIS hiperspektral görüntülerinin sınıflandırılması ve tematik harita üretiminde RotOr, KKO ve RO sınıflandırıcıları değerlendirmeye alınmıştır. Sınıflandırma modellerinin oluşturularak kontrollü sınıflandırma işleminin gerçekleştirilmesi amacıyla her iki görüntü için mevcut bulunan yer doğrulama verileri içerisinde her bir sınıf için toplam örnek sayısının %10'u eğitim veri seti için kullanılırken, geri kalan pikseller oluşturulan sınıflandırma modellerinin performans analizinde kullanılmak üzere test veri seti olarak kullanılmıştır (Çizelge 1). RotOr, KKO ve RO algoritmaları ile toplu öğrenme modeli veya karar ormanı oluşumunda

değerlendirmeye alınacak ağaç sayısı kullanıcı tarafından belirlenen bir parametredir. KKO algoritmasının en önemli özelliklerinden birisi kullanılacak ağaç sayısının algoritmanın sınıflandırma performansında anlamlı değişiklikler ortaya çıkarmaması olarak ifade edilmektedir. Bu KKO algoritması için ağaç sayısı standart olarak 200 olarak önerilmiş ve uygulamada her iki görüntü için bu ağaç sayısı esas alınmıştır. RO ve RotOr algoritmaları için ağaç sayısı 5-katlı çapraz geçerlilik yaklaşımıyla belirlenmiştir. Sonuç olarak RotOr için ağaç sayısı AVIRIS görüntüsü için 200, ROSIS görüntüsü için 100 olarak belirlenirken; RO algoritması için ağaç sayısı AVIRIS görüntüsü için 200, ROSIS görüntüsü için 500 olarak tespit edilmiştir.

Çizelge 1. Indian Pine (AVIRIS) ve Pavia Üniversitesi (ROSI) hiperspektral uydu görüntülerine ilişkin arazi örtüsü sınıfları ve toplam örnek piksel sayıları.

<i>Indian Pine (AVIRIS) görüntüsü</i>			<i>Pavia Üniversitesi (ROSI) görüntüsü</i>		
Sınıf No	Arazi örtüsü sınıfı	Piksel sayısı	Sınıf No	Arazi örtüsü sınıfı	Piksel sayısı
1	Alfaalfa (Yonca)	54	1	Asphalt (Asfalt)	6631
2	Corn-notill (Mısır-işlenmemiş toprak)	1434	2	Meadow (Çim)	18649
3	Corn-mintill (Mısır- az işlenmiş toprak)	834	3	Gravel (Çakıl)	2099
4	Corn (Mısır)	234	4	Tree (Ağaç)	3064
5	Grass-pasture (Çayır-otlak)	497	5	Metal Sheet (Metal sac)	1345
6	Grass-trees (Çayır-ağaç)	747	6	Bare Soil (Bozkır)	5029
7	Grass-pasture-mowed (Biçilmiş-çayır-otlak)	26	7	Bitumen (Zift)	1330
8	Hay-windrowed (Saman)	489	8	Brick (Tuğla)	3682
9	Oats (Yulaf)	20	9	Shadow (Gölge)	947
10	Soybean-notill (Soya fasulyesi -işlenmemiş toprak)	968			
11	Soybean-mintill (Soya fasulyesi- az işlenmiş toprak)	2468			
12	Soybean-clean (Soya fasulyesi)	614			
13	Wheat (Buğday)	212			
14	Woods (Orman)	1294			
15	Buildings-Grass-Trees-Drives (Binalar-çayır-ağaç)	380			
16	Stone-Steel-Towers (Taş-çelik-kuleler)	95			
Toplam		10.366	Toplam		42.776

Toplu öğrenme algoritmaları için tespit edilen optimum parametre değerleri ile AVIRIS ve ROSIS görüntüleri için sınıflandırma modelleri oluşturulmuş ve RotOr, KKO ve RO algoritmalarının test veri seti için hesaplanan genel sınıflandırma doğrulukları, Kappa değerleri ve ölçülen işlem zamanı Çizelge 2’de verilmiştir. Çizelgede her iki uydu görüntüsü için hesaplanan en yüksek sınıflandırma doğrulukları, Kappa değerleri ve en düşük işlem süresi koyu olarak gösterilmiştir. Çizelgeden de görüleceği üzere özellikle tarımsal ürünlerin spektral olarak büyük benzerlikler gösterdiği 16 arazi örtüsü sınıfını içeren AVIRIS görüntüsünün sınıflandırılmasında KKO algoritması ile hesaplanan genel doğruluklar ve Kappa değerleri RotOr algoritmasına göre %5, RO algoritmasına göre %12 daha yüksektir.

Çizelge 2. RotOr, KKO ve RO algoritması için hesaplanan doğruluk ölçütleri.

Algoritma	<i>Indian Pine (AVIRIS) görüntüsü</i>			<i>Pavia Üniversitesi (ROSI) görüntüsü</i>		
	Genel doğruluk (%)	Kappa değeri	İşlem süresi (saniye)	Genel doğruluk (%)	Kappa değeri	İşlem süresi (saniye)
RotOr	82,27	0,796	21,58	92,85	0,908	30,34
KKO	87,13	0,852	26,76	94,32	0,924	24,16
RO	74,78	0,709	3,31	89,35	0,856	19,88

Diğer taraftan farklı yapı malzemeleri ile inşa edilen yapılaşmış alanların yoğun olduğu ROSIS görüntüsünün sınıflandırılması neticesinde KKO algoritması kullanımı ile RotOr algoritmasına göre %2, RO algoritmasına göre %5 daha doğru sınıflandırma sonuçları elde edilmiştir. Sınıflandırma doğrulukları arasındaki söz konusu farklılıkların McNemar’s testi sonuçlarına göre istatistiksel olarak anlamlı olduğu görülmüştür. Elde edilen bu sonuçlar özellikle uygulamada dikkate alınan karmaşık yapı hiperspektral verilerin sınıflandırılmasında kanonik korelasyon ormanı algoritmasının daha başarılı sonuçlar verdiğini göstermektedir. Toplu öğrenme modeli oluşumunda ihtiyaç duyulan işlem süresi açısından sonuçlar ele alındığında RO orman algoritmasının RotOr ve KKO algoritmalarına göre çok daha hızlı olduğu görülmektedir.

Karar ormanı gibi tek bir sınıflandırıcının kullanıldığı toplu öğrenme algoritmalarının sınıflandırma işlemindeki başarısı, ormandaki her bir ağacın farklı eğitim verileri kullanılarak oluşturulması veya karar ormanı içerisindeki çeşitliliğin sağlanmasıyla doğrudan ilişkilidir (Diez-Pastor vd. 2015). Bu nedenle uygulama kapsamında RotOr, KKO ve RO algoritmalarına ilişkin hesaplanan sınıflandırma doğrulukları ve karar ormanındaki çeşitlilik arasındaki ilişkinin ortaya koyulması amacıyla Kuncheva ve Whitaker (2003) tarafından önerilen 3 farklı çeşitlilik ölçütü (Q-istatistiği, korelasyon katsayısı ve entropi) değerlendirmeye alınmıştır. Q-istatistiği (Q) iki sınıflandırma sonucu arasındaki benzerliği ölçerken, korelasyon katsayısı (ρ) ise iki sınıflandırma sonucu arasındaki korelasyonu ölçmektedir. Dolayısıyla düşük Q ve ρ değerleri çeşitliliğin daha yüksek olduğunu ifade etmektedir. Entropi (Ent) ölçütü 0 ile 1 arasında değerler alırken, 0 sınıflandırma sonuçları arasında farklılık olmadığını, 1 farklılık olduğunu göstermektedir. Dolayısıyla yüksek Ent değerleri çeşitliliğin yüksek olduğunu ifade etmektedir. RotOr, KKO ve RO için hesaplanan istatistiki değerler Çizelge 3'te verilmiştir. Çizelgeden de görüleceği üzere her iki hiperspektral görüntü içinde en yüksek entropi değeri ile en düşük Q ve ρ değerlerine KKO algoritması ile ulaşılmıştır. Elde edilen bu sonuç, hesaplanan sınıflandırma doğruluklarına paralel olup, karar ormanındaki çeşitliliğin sınıflandırma doğruluğu ile ilişkili olduğunu destekler niteliktedir.

Çizelge 3. RotOr, KKO ve RO algoritması için hesaplanan çeşitlilik ölçütlerine ilişkin istatistiki sonuçlar.

Algoritma	Indian Pine (AVIRIS) görüntüsü			Pavia Üniversitesi (ROSIS) görüntüsü		
	Q (↓)	ρ (↓)	Ent (↑)	Q (↓)	ρ (↓)	Ent (↑)
RotOr	0,5609	0,2979	0,4903	0,6851	0,2838	0,2501
KKO	0,5074	0,2573	0,5061	0,6728	0,2731	0,2927
RO	0,6054	0,3324	0,4779	0,7140	0,3387	0,2483

AVIRIS ve ROSIS hiperspektral görüntüleri için oluşturulan RotOr, KKO ve RO sınıflandırma modelleri her iki görüntüye uygulanarak tematik haritalar üretilmiştir. Indian Pine (AVIRIS) görüntüsü için RotOr, KKO ve RO toplu öğrenme modelleri ile üretilen tematik haritalar Şekil 6'da verilmiştir. Şekilden de görüleceği üzere RO algoritması ile üretilen tematik haritada (Şekil 6c) 16 sınıfı temsil eden pikseller arasındaki hatalı sınıflandırmaların çok fazla olduğu dolayısıyla tematik harita doğruluğunun diğer algoritmalarla göre düşük olduğu açıkça görülmektedir. Diğer taraftan, tematik haritaların görsel analizi neticesinde KKO ile üretilen tematik haritada (Şekil 6b) sınıflar arasındaki ayrımın daha kolay olduğu ifade edilebilir. Özellikle sınırlı sayıda yer doğrulama verisine sahip yulaf (20 piksel), biçilmiş-çayır-otlak (26 piksel) ve yonca (54 piksel) sınıfları için KKO algoritması ile daha doğru bir sınıflandırma yapıldığı görülmektedir.

Şekil 6. Indian Pine (AVIRIS) görüntüsü için a) RotOr, b) KKO ve c) RO ile üretilen tematik haritalar.

Pavia Üniversitesi (ROSIS) görüntüsü için RotOr, KKO ve RO toplu öğrenme modelleri ile üretilen tematik haritalar Şekil 7'de verilmiştir. Şekilden de görüleceği RotOr ve KKO algoritmaları yardımıyla üretilen tematik haritalar ROSIS görüntüsü için mevcut 9 arazi örtüsü sınıfının temsiliyeti noktasında daha başarılıdır. Özellikle su geçirimsiz yüzeyler olarak bilinen metal sac, tuğla, asfalt, çakıl ve zift sınıflarını temsil eden pikseller söz konusu haritalarda daha doğru bir şekilde sınıflandırılmıştır. Diğer taraftan RO algoritması ile üretilen tematik haritada benzer spektral özelliklere sahip çim, ağaç ve bozkır sınıflarını temsil eden pikseller arasındaki hatalı sınıflandırmaların daha fazla olduğu görülmektedir.

Şekil 7. Pavia Üniversitesi (ROSIS) görüntüsü için a) RotOr, b) KKO ve c) RO ile üretilen tematik haritalar.

[1]

5. SONUÇ

Bu çalışmada, karar ormanı olarak bilinen özellikle son yıllarda yaygın olarak kullanılan rastgele orman (RO) algoritmasının performansını arttırmak üzere geliştirilen rotasyon orman (RotOr) ve kanonik korelasyon orman (KKO) algoritmalarının hiperspektral görüntülerin sınıflandırılmasındaki performansları istatistiksel yaklaşımlarla detaylı bir şekilde incelenmiştir. Hiperspektral görüntülerin sınıflandırılmasında karar ormanı algoritmalarının kullanılmasına ilişkin elde edilen sınıflandırma sonuçlarından bazı önemli bulgular elde edilebilir. Sınıflandırma performansları açısından sonuçlar ele alındığında, özellikle benzer spektral özelliklere sahip 16 arazi örtüsü sınıfı için sınırlı sayıda yer doğrulama verisi içeren Indian Pine (AVIRIS) görüntüsünün sınıflandırılmasında KKO algoritması kullanımıyla sınıflandırma doğruluğunda RotOr algoritmasına göre %5, RO algoritmasına göre %12'ye varan seviyelerde artış olduğu görülmektedir. Diğer taraftan, Pavia Üniversitesi (ROSIS) görüntüsüne ilişkin sınıflandırma sonuçları KKO algoritması ile ağırlıklı olarak su geçirimsiz yüzeylerden oluşan 9 arazi örtüsü sınıfını temsil eden piksellerin sınıflandırılmasında KKO algoritmasının RotOr algoritmasına göre %2, RO algoritmasına göre %5 düzeyinde daha doğru sonuçlar ürettiğini göstermektedir. Her iki görüntü için sınıflandırma doğruluğunda bu düzeyindeki söz konusu iyileşmeler McNemar's test sonuçlarına göre istatistiksel olarak anlamlı olup, KKO algoritmasının sınıflandırma işlemindeki üstünlüğünü destekler niteliktedir. Bu durum kanonik korelasyon analizinde sadece bantlar arasındaki korelasyona değil bantlar ile sınıflar arasındaki korelasyonunda dikkate alınmasının bir sonucu olduğu düşünülmektedir. Sınıflandırma işlemi için gerekli işlem süresi açısından sonuçlar ele alındığında, RO algoritması diğer algoritmalarla göre oldukça az işlem süresi gerektirdiği tespit edilmiş olup, algoritmanın problem çözmedeki hızı literatürdeki birçok çalışmayı destekler niteliktedir. Karar ağaçlarını esas alan homojen yapılı toplu öğrenme algoritmalarının performansı çeşitlilik (diversity) ve doğruluk arasındaki ilişki ele alındığında, KKO algoritmasının her iki görüntü için yüksek doğrulukta sonuçlar ürettiği ve aynı zamanda çalışmada ele alınan çeşitlilik ölçütlerine göre (Q-istatistiği, korelasyon katsayısı ve entropi) ormandaki ağaçların çeşitliliği noktasında daha başarılı olduğu tespit edilmiştir. Elde edilen bu sonuç, karar ormanı olarak bilinen toplu öğrenme yaklaşımlarının kullanımında ormandaki çeşitliliğin sınıflandırma doğruluğu ile yakın ilişkili olduğu göstermektedir. Tüm sonuçlar bir arada ele alındığında, spektral olarak benzer özelliklere sahip arazi örtüsü sınıflarını içeren hiperspektral özellikteki görüntülerin sınıflandırılması gibi karmaşık yapılı sınıflandırma problemlerinin çözümünde KKO ve RotOr olarak adlandırılan güncel karar ormanı algoritmalarının daha başarılı olduğunu göstermektedir.

KAYNAKLAR

- Belgiu, M., Drăguț, L.,** 2016. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31.
- Breiman, L.,** 2001. Random forests. *Machine Learning*, 45(1), 5-32.
- Chan, J. C. W., Paelinckx, D.,** 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112(6), 2999-3011.

- Çölkesen, İ., 2017. Fonksiyonel Ağaçlar ile Obje-Tabanlı Sınıflandırma: WorldView-2 Uydu Görüntüsü Örneği", *Harita Dergisi*, 157(1),9-21.
- Çölkesen, İ., Kavzoglu, T., 2017. Kanonik Korelasyon Orman Algoritması ile Uzaktan Algılanmış Görüntülerin Sınıflandırılması, *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 17(1) Özel Sayı, 102-111.
- 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-Orsio, C. I., Kuncheva, L. I., 2015. Diversity Techniques Improve the Performance of the Best Imbalance Learning Ensembles. *Information Sciences*, 325, 98-117.
- Ghamisi, P., Plaza, J., Chen, Y. S., Li, J., Plaza, A., 2017. Advanced Spectral Classifiers for Hyperspectral Images A review. *IEEE Geoscience and Remote Sensing Magazine*, 5(1), 8-32.
- Kavzoglu, T., Colkesen, I., 2013. An Assessment of Effectiveness of Rotation Forest Ensemble for Land Use and Land Cover Mapping, *International Journal of Remote Sensing*, 34(12), 4224-4241.
- Kuncheva, L.I., Whitaker, C. J., 2003, Measures of Diversity in Classifier Ensembles and Their Relationship with The Ensemble Accuracy. *Machine Learning*, 51(2), 181-207.
- 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, 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), 823-870.
- Maxwell, A. E., Warner, T. A., Fang, F., 2018. Implementation of Machine-learning Classification in Remote Sensing: An Applied Review. *International Journal of Remote Sensing*, 39(9), 2784-2817.
- Rainforth, T., Wood, F., 2015. Canonical Correlation Forests. <https://arxiv.org/pdf/1507.05444.pdf> (Erişim zamanı 21.06.2018).
- Ren, Y., Zhang, L., Suganthan, P. N., 2016. Ensemble Classification and Regression-Recent Developments, Applications and Future Directions. *IEEE Computational Intelligence Magazine*, 11(1), 41-53.
- Rodríguez, J. J., Kuncheva, L. I., 2006. Rotation Forest: A New Classifier Ensemble Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10), 1619-1630.
- Rokach, L., 2010, Ensemble-based Classifiers. *Artificial Intelligence Review*, 33(1-2), 1-39.
- Rokach, L., 2016, Decision Forest: Twenty Years of Research. *Information Fusion*, 27, 111-125.
- Sagi, O., Rokach, L., 2018. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, 8(4).
- Tso, B., Mather, P. M., 2009. *Classification Methods for Remotely Sensed Data*, 2nd, London: Taylor & Francis.
- Xia, J. S., Du, P. J., He, X. Y., Chanussot, J., 2014. Hyperspectral Remote Sensing Image Classification Based on Rotation Forest. *IEEE Geoscience and Remote Sensing Letters*, 11(1), 239-243.
- Zhang, H. S., Wang, T., Liu, M. F., Jia, M. M., Lin, H., Chu, L. M., Devlin, A. T., 2018. Potential of Combining Optical and Dual Polarimetric SAR Data for Improving Mangrove Species Discrimination Using Rotation Forest. *Remote Sensing*, 10(3).