

ROTASYON ORMAN ALGORİTMASI İLE YÜKSEK ÇÖZÜNÜRLÜKLÜ MULTİSPEKTRAL UYDU GÖRÜNTÜLERİNİN SINIFLANDIRILMASI

İsmail ÇÖLKESEN¹, Tahsin YOMRALIOĞLU², Taşkın KAVZOĞLU³

¹Araş. Gör., Gebze Yüksek Teknoloji Enstitüsü, Jeodezi ve Fotogrametri Mühendisliği Bölümü, 41400, Gebze, Kocaeli, icolkesen@gyte.edu.tr

²Prof. Dr., İstanbul Teknik Üniversitesi, Geomatik Mühendisliği Bölümü, 34469, Maslak, İstanbul, tahsin@itu.edu.tr

³Prof. Dr., Gebze Yüksek Teknoloji Enstitüsü, Jeodezi ve Fotogrametri Mühendisliği Bölümü, 41400, Gebze, Kocaeli, kavzoglu@gyte.edu.tr

ÖZET

Çok yüksek çözünürlüklü uydu görüntüleme sensörlerinin kullanılabilirliğinin artması arazi kullanımı ve arazi örtüsü türleri hakkında ayrıntılı bilgi elde edilmesi noktasında yeni bir veri kaynağı sağlamıştır. Korelasyonlu ve spektral olarak benzer karakteristiklere sahip pikselleri içeren yeni nesil yüksek çözünürlüklü uydu görüntüleri üzerinden doğruluğu yüksek tematik harita üretimi için literatürde çeşitli gelişmiş sınıflandırma teknikleri ve sınıflandırıcılar önerilmiştir. Birden çok sınıflandırıcının tahminlerinin birleştirilmesi esasını temel alan toplu öğrenme algoritmaları yüksek çözünürlüklü uydu görüntülerinin sınıflandırılmasında yaygın bir kullanıma sahiptir. Rotasyon orman (RotOR) algoritması son yıllarda sınıflandırıcıların performansının artırılması amacıyla önerilen yeni nesil bir toplu öğrenme algoritmasıdır. Bu çalışmada, RotFOR algoritmasının arazi örtüsü ve arazi kullanımının haritalanmasındaki sınıflandırma performansının araştırılmasında temel veri kaynağı olarak sekiz spektral banda sahip yüksek çözünürlüklü WorldView-2 görüntüsü kullanılmıştır. Yöntemin sınıflandırma performansı hızlandırma ve rastgele orman olarak bilinen iki popüler toplu öğrenme algoritması ile karşılaştırılmış ve istatistiksel olarak analiz edilmiştir. Bununla birlikte toplu öğrenme algoritmalarının etkinliğinin araştırılması amacıyla en çok benzerlik algoritması da sınıflandırma işleminde değerlendirilmeye alınmıştır. Elde edilen sonuçlar hem sınıflandırma doğrulukları hem de McNemar test istatistiği sonuçlarına göre RotOR algoritmasının diğer toplu öğrenme algoritmalarından daha üstün olduğunu göstermektedir. Sınıflandırma performansları karşılaştırıldığında RotOR algoritması ile %3'e varan yüksek sınıflandırma doğrulukları elde edildiği ve sınıflandırma performansları arasındaki farklılıkların istatistiksel olarak anlamlı olduğu görülmüştür. Sonuç olarak, bu çalışma RotOR algoritmasının çok yüksek çözünürlüklü uydu görüntülerinin sınıflandırılmasındaki etkinliğini ortaya koymaktadır.

Anahtar Sözcükler: Rotasyon Orman, Hızlandırma, Rastgele Orman, WorldView-2, McNemar test

ABSTRACT

CLASSIFICATION OF MULTISPECTRAL HIGH-RESOLUTION SATELLITE IMAGERY USING ROTATION FOREST ALGORITHM

Increasing availability of very high resolution satellite imaging sensors provides a new data source for gathering detailed information about land use and land cover types of the Earth's surface. Various advanced classification techniques and classifiers have been suggested in the literature to produce accurate thematic maps from the new generation of high resolution satellite imageries containing pixels that of correlated and spectrally similar characteristics. Ensemble learning based on the strategies that the combine the prediction of individual classifiers has been most commonly applied for the classification of the high resolution imageries. A novel ensemble learning algorithm called rotation forest (RotFor) has been recently suggested to improve the individual classifier performances. In this study, WorldView-2 imagery having high spatial resolution with eight spectral bands was used as a main data source to investigate classification performance of RotFOR algorithm in land use and land cover mapping. The classification performance was compared and statistically analyzed with two popular ensemble learning algorithm namely, boosting and random forest. In addition, maximum likelihood algorithm was also utilized for the classification task to explore the effectiveness of the ensemble learning algorithms. Results showed that RotFor algorithm outperformed the other ensemble learning algorithms in terms of both classification accuracy and McNemar's test statistics. When the classification performances were compared, it was observed that higher classification accuracies up to 3% obtained with RotFor algorithm and the differences in the classification performances were found to be statistically significant. To sum up, this study verifies the effectiveness of RotFor in the classification of very high resolution satellite imagery.

Keywords: Rotation Forest, Boosting, Random Forest, WorldView-2, McNemar's test.

1. GİRİŞ

Uydu görüntülerinin sınıflandırılması yardımıyla yeryüzünün farklı özelliklerini temsil eden tematik harita üretimi uzaktan algılama alanında önemli bir çalışma konusudur. Özellikle son yıllarda uydu teknolojileri ve algılama sensörlerinin tasarımında yaşanan gelişmelere paralel olarak yüksek konumsal ve spektral çözünürlüğe sahip uydu görüntüleri elde edilebilmektedir. Yeni nesil uydu görüntüleri kullanıcılara yeryüzü nesnelere ilişkin gerek konumsal gerekse spektral açıdan daha detaylı bilgiler sağlamaktadır. Yaşanan teknolojik gelişmeler beraberinde uzaktan algılama alanında yeni araştırma konularının ortaya çıkmasına neden olmuştur. Kentsel alanlara yönelik gerçekleştirilen çalışmalarda şehirleşmenin izlenmesi, bina, betonarme yapılar ve yol gibi su geçirimsiz yüzeylerin haritalanması öne çıkan araştırma konularındandır (Patino et al., 2013; Huang et al., 2014; Mei et al., 2014).

Bununla birlikte farklı ağaç ve bitki türlerinin ayırt edilmesi, ekonomik değeri olan ağaç türlerinin tespiti ve ağaç sayımları, biokütle hesabı, ağaç canlılığının izlenmesi ve rekolte tahminleri gibi ekolojik çevre açısından önemli bir çok uygulama da yüksek çözünürlüklü uydu görüntülerinden yararlanılmaktadır (Garrity et al., 2013; Waser et al., 2014; Singh et al., 2014).

Yüksek çözünürlüklü uydu görüntülerinin kullanılması birçok doğal ve yapay nesnenin bir arada değerlendirilerek benzer spektral özelliklere sahip nesnelerin birbirinden ayırt edilmesini gerektirmektedir. Diğer bir ifadeyle yüksek çözünürlüklü uydu görüntüleri yardımıyla arazi örtüsü/kullanımının belirlenmesinde benzer spektral özelliklere sahip karmaşık yapıdaki piksellerin değerlendirilmesi ve sınıflandırılması söz konusudur. Literatürde sınıflandırma işleminin gerçekleştirilmesi ve tematik harita üretimine yönelik birçok algoritma önerilmiştir (Lu and Weng, 2007). Bunlar arasında hesaplama açısından hızlı ve matematiksel olarak kolay anlaşılabilen en çok benzerlik yöntemi yaygın bir kullanıma sahiptir. Yöntem uygulama açısından pratik olmasına rağmen görüntü üzerindeki piksellerin normal dağılımda olduğu kabulünü esas almaktadır. Bu nedenle yöntem birbirine benzer spektral özelliğe sahip karmaşık ve homojen olmayan yapıdaki piksellerin ayırt edilmesi noktasında yetersiz kalmaktadır. Bu problemin çözümüne yönelik olarak son yıllarda ileri sınıflandırma teknikleri kullanılmaya başlanmıştır. Söz konusu algoritmalar içerisinde tek bir sınıflandırıcı yerine birden çok sınıflandırıcının değerlendirmeye alınması ve sınıflandırıcılardan elde edilen tahminlerden yola çıkılarak sınıflandırma işleminin gerçekleştirilmesi esasına dayalı toplu öğrenme algoritmaları en çok tercih edilenler arasındadır. Özellikle yüksek çözünürlüklü uydu görüntülerinin sınıflandırılması ve tematik harita üretiminde hızlandırma, torbalama ve rastgele orman olarak bilinen toplu öğrenme algoritmaları başarıyla kullanılmaktadır (Malinverni et al., 2011; Demarchi et al., 2014; Hayes et al., 2014; Ghosh and Joshi, 2014). Son yıllarda ana bileşenler analizi kullanarak veri setinin yeniden oluşturulması prensibini esas alan rotasyon orman algoritması uzaktan algılanmış görüntülerin sınıflandırılmasında önerilen fakat sınırlı sayıda test edilmiş yeni nesil toplu öğrenme algoritmasıdır (Rodriguez et al., 2006; Kavzoglu and Colkesen, 2013).

Bu çalışmada rotasyon orman algoritmasıyla 2013 tarihinde kaydedilen WorldView-2 görüntüsünün sınıflandırılması ve çalışma alanına ait arazi örtüsü/kullanımını gösteren tematik harita üretilmesi amaçlanmıştır. Algoritmanın sınıflandırma performansı popüler toplu öğrenme algoritmalarından rastgele orman ve hızlandırma algoritmaları ile karşılaştırılmıştır. Bununla birlikte toplu öğrenme algoritmalarının yüksek çözünürlüklü görüntülerin sınıflandırılmasındaki performansı geleneksel yöntem olarak kabul edilen en çok benzerlik sınıflandırıcısı ile karşılaştırılarak analiz edilmiştir. Elde edilen sınıflandırma doğruluklarının karşılaştırılmasının yanında doğruluklar arasındaki farklılıkların istatistiksel anlamlılığı McNemar testi kullanılarak analiz edilmiştir.

2. ÇALIŞMA ALANI VE KULLANILAN VERİ

Rotasyon orman (RotOR) algoritmasının yüksek çözünürlüklü uydu görüntüsü üzerinden karmaşık yapıdaki arazi örtüsü/kullanım türlerinin sınıflandırılmasındaki performansının ortaya koyulması amacıyla Gebze Yüksek Teknoloji Enstitüsü kampüs alanını da içerisine alan yaklaşık 300 hektarlık bir bölge çalışma alanı olarak belirlenmiştir (Şekil 1). Çalışma alanı içerisinde temel arazi örtüsü olarak bilinen bozkır, su ve toprak alanlarına ilave olarak benzer spektral özelliklere sahip geniş ve iğne yapraklı ağaç türleri, su geçirimsiz yüzey olarak adlandırılan farklı yapıdaki bina türleri ve asfalt zeminli yollar mevcuttur. Çalışma alanında yapılan detaylı arazi çalışmaları sonucunda bölgede 13 temel arazi örtüsü/kullanım sınıfı olduğuna karar verilmiştir. Arazi örtüsü olarak bozkır, su ve toprak sınıflarının yanında geniş ve iğne yapraklı ağaç türlerini temsil eden 5 farklı arazi örtüsü sınıfı tanımlanmıştır. Bu sınıflar geniş yapraklı ağaç türlerini temsilen çınar ve zeytin sınıfları, iğne yapraklı ağaç türlerini temsil eden fıstık çamı, kızılçam ve sedir sınıfları olarak adlandırılmıştır. Bölgenin arazi kullanım sınıfları olarak yapılaşmış alanlar veya geçirimsiz yüzeyleri temsil eden 4 farklı arazi kullanım sınıfı tanımlanmıştır. Söz konusu arazi kullanım sınıfları kırmızı, beyaz ve gri çatıya sahip olan binalar ve bölgeyi çevreleyen ara ve ana yolları temsil eden yol sınıfıdır. Belirlenen arazi örtüsü/kullanım sınıflarına ilave olarak yüksek çözünürlüklü uydu görüntülerinde söz konusu olan gölge problemini ve gölge probleminden kaynaklanabilecek sınıflandırma hatalarının azaltılması amacıyla gölge sınıfı tanımlanmıştır.

Belirlenen 13 farklı arazi örtüsü/kullanım türlerinin RotOR algoritmasıyla sınıflandırılması için çalışma alanını kapsayan 2013 tarihinde kaydedilmiş WorldView-2 (WV2) uydu görüntüsü kullanılmıştır. WV2 görüntüsünün kırmızı (630-690 nm), yeşil (510-580 nm), mavi (450-510 nm) ve yakın kızılötesi (770-895 nm) bantlarına ilave olarak kıyı (400-450nm), sarı (585-625 nm), kırmızı-kenar (705-745 nm) ve ikinci kızılötesi (860-1040 nm) bantları uygulama da değerlendirmeye alınmıştır. Uydu görüntüsünün geometrik düzeltilmesi ve UTM koordinat sistemine dönüşümünde mevcut hâlihazır haritalar ve bölgeye ait yüksek çözünürlüklü görüntülerden yararlanılmıştır. WV-2 görüntüsünün sahip olduğu 0,5m konumsal çözünürlüğe sahip pankromatik bant ile 2m konumsal çözünürlüklü multispektral bantlarının kaynaştırılmasında Gram-Schmidt algoritması kullanılmıştır. Kaynaştırma sırasında 3x3 boyutlu yumuşatma filtresi uygulanmış ve yeniden örnekleme işleminde bilinear enterpolasyon algoritması kullanılmıştır.



Şekil 1. Çalışma alanının konumu ve uygulamada kullanılan WorldView-2 görüntüsü.

3. SINIFLANDIRMA ALGORİTMALARI

Uzaktan algılanmış görüntüler yardımıyla yeryüzü objelerine ilişkin bilgi çıkarımında en sık kullanılan yöntem görüntü sınıflamasıdır. Sınıflandırma işlemi görüntü üzerindeki her bir piksele ait yansıma ve parlaklık değerleri kullanılarak piksellerin kullanıcı tarafından belirlenen sınıflara ayrılması işlemidir. Sınıflandırma sonucunda farklı arazi örtüsü ve kullanımını temsil eden tematik haritalar üretilmekte ve elde edilen haritalar yeryüzü özelliklerinin yorumlanması ve analizinde kullanılmaktadır. Bu çalışmada WorldView-2 uydu görüntüsünün sınıflandırılması ve 13 arazi örtüsü/kullanım sınıfını temsil eden tematik harita üretiminde toplu öğrenme algoritmalarından hızlandırma, rastgele orman ve rotasyon orman algoritmaları kullanılmıştır. Sınıflandırma performanslarının karşılaştırılması amacıyla geleneksel sınıflandırma yöntemi olarak bilinen en çok benzerlik algoritması da sınıflandırmada kullanılmıştır.

Kullanılan algoritmaların performanslarının değerlendirilmesinde öncelikli olarak sınıflandırma sonucunda elde edilen genel doğruluklar dikkate alınmıştır. Buna ilaveten hesaplanan sınıflandırma doğrulukları arasındaki farklarının istatistiksel anlamlılığı McNemar testi kullanılarak analiz edilmiştir. Ki-kare dağılımını esas alan McNemar test istatistiği Eşitlik 1 yardımıyla hesaplanmaktadır. Eşitlikteki değişkenler 2×2 boyutlu bir hata matrisinin elemanları olup, n_{ij} , i . sınıflandırıcı tarafından hatalı j . sınıflandırıcı tarafından doğru sınıflandırılan piksel sayısını, n_{ji} ise j . sınıflandırıcı tarafından hatalı i . sınıflandırıcı tarafından doğru sınıflandırılan piksel sayısını ifade etmektedir (Foody, 2004). Hesaplanan istatistik değeri Ki-kare tablo değerinden (%95 güven aralığında $\chi^2 = 3,84$) büyük olduğunda değerlendirmeye alınan iki sınıflandırma doğruluğu arasındaki farkın istatistiksel olarak anlamlı olduğu söylenebilir. Diğer bir ifadeyle, elde edilen bu istatistiksel sonuç iki sınıflandırma doğruluğunun birbirinden farklı olduğunu göstermektedir.

$$McNemar (\chi^2) = \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} \quad (1)$$

3.1. En Çok Benzerlik Algoritması

En çok benzerlik (EÇB) algoritması uzaktan algılanmış görüntülerin sınıflandırılmasında yaygın olarak kullanılan parametrik bir kontrollü sınıflandırma algoritmasıdır. EÇB algoritması sınıflandırma işleminin gerçekleştirilmesinde ortalama değer, varyans ve kovaryans gibi istatistiksel değerleri dikkate almakta ve sınıfları oluşturan piksellerin normal dağılımda olduğu kabulünü esas aldığından EÇB parametrik bir algoritma olarak tanımlanmaktadır. EÇB

algoritması ile bir pikselin p sayıdaki bant değerlerini içeren x vektörünün bir k sınıfta olma olasılığı Eşitlik 2 yardımıyla hesaplanır (Mather, 1999). Eşitlikte; $P(x)$, olasılık değeri; S_i , i . sınıf için varyans-kovaryans matrisini; $| \cdot |$, matrisin determinantını ifade etmektedir.

$$P(x)=2\pi^{-0.5p} |S_i|^{-0.5} \exp \left[-0.5 \left(y^T S_i^{-1} y \right) \right] \quad (2)$$

3.2. Toplu Öğrenme Algoritmaları

Toplu öğrenme algoritmaları tek bir sınıflandırıcı yerine birden çok sınıflandırıcıyı içerecek şekilde bir sınıflandırıcı kümesinin oluşturulması ve küme içerisindeki sınıflandırıcıların tahminlerinden yola çıkılarak sınıf etiketi belli olmayan piksellerin sınıflandırılması esasına dayanmaktadır (Dietterich, 2000a). Toplu öğrenme veya sınıflandırıcıların birleştirilmesi, yeniden örneklenen eğitim setleri ile sınıflandırıcıların ayrı ayrı eğitilmesi ve sonuçta ortaya çıkan tahminler ile sınıflandırma işleminin gerçekleştirilmesi işlemlerini içermektedir. Literatürde toplu öğrenme algoritmaları kullanımı ile sınıflandırma performansında önemli derecede artış olduğu vurgulanmaktadır (Opitz and Maclin, 1999; Dietterich, 2000b, Pal and Mather, 2003). Günümüze kadar toplu öğrenme algoritması olarak başta hızlandırma, torbalama ve rastgele orman olmak üzere birçok algoritma geliştirilmiştir (Kavzoglu and Colkesen, 2013).

Hızlandırma algoritmasının temeli bir dizi sınıflandırıcı serisi oluşturulmasına dayanmaktadır (Freund and Schapire, 1996; Schapire, 1999; Nanni and Lumini, 2006; Çölkesen ve Kavzoğlu, 2009). Serinin üyesi olan her bir sınıflandırıcı için kullanılan eğitim seti, serideki bir önceki sınıflandırıcı veya sınıflandırıcıların performansına göre belirlenir. Bu amaçla bir önceki sınıflandırıcı tarafından hatalı tahmin edilen pikseller bir sonraki sınıflandırıcı için oluşturulacak eğitim setinde doğru tahmin edilen piksellere göre daha fazla tekrar edilir. Diğer bir ifadeyle hatalı sınıflandırılan piksellerin eğitim veri seti içerisindeki ağırlıkları artırılır. Hızlandırma algoritması bu iteratif yaklaşımla eğitim veri seti içerisinde doğru sınıflandırılan pikseller yerine hatalı sınıflandırılan piksellere yoğunlaşarak sınıflandırma doğruluğunu artırılmasını amaçlamaktadır. Popüler toplu öğrenme algoritması olarak bilinen rastgele orman (RO) algoritması temel sınıflandırıcı olarak karar ağaçlarını kullanmaktadır. Temelde bir çeşit karar ağacı algoritması olan RO algoritması eğitim aşamasında birden çok karar ağacını kullanmakta ve bu yapısıyla karar ağacı ormanı olarak tanımlanabilmektedir (Breiman, 2001). Eğitim işlemi sırasında ormandaki her bir karar ağacı için eğitim veri seti içerisinde rastgele alt kümeler oluşturulur. Söz konusu alt kümelerin 2/3'ü karar ağacı yapısını oluşturmak için kullanılırken, diğer kısmı ağaç yapısının geçerliliğini test etmek için kullanılır. Ormandaki her bir karar ağacı sınıflandırma sonucu bir oy alır ve orman içerisindeki tüm ağaçlardan en çok oy alan (en düşük hata oranına sahip) belirlenerek sınıflandırmaya esas ağaç yapısı tespit edilir. Sınıf etiketi bilinmeyen herhangi bir örnek (piksel), tüm ağaç tahminlerinde en fazla oy alan sınıfa atanması suretiyle sınıflandırılır.

Son yıllarda literatürde kullanılmaya başlanan yeni nesil topluluk öğrenme algoritması olarak ortaya koyulan rotasyon orman (RotOR) algoritması, bir özellik çıkarma tekniği (ana bileşenler analizi) kullanarak sınıflandırıcı topluluğunun oluşturulmasını esasına dayanmaktadır (Rodriguez et al., 2006; Kavzoglu and Colkesen, 2013). RotOR algoritmasının temel çalışma prensibi rastgele orman algoritmasına benzer şekilde olup birden fazla ağaç kullanılmaktadır. Ancak ormandaki her bir karar ağacının eğitiminde kullanılacak veri seti ana bileşen analizi yardımıyla belirlenmektedir. RotOR algoritması ile ormandaki karar ağaçlarının eğitimi aşamasında eğitim veri seti rastgele alt kümelere bölünür ve her bir alt küme ana bileşenler analizi uygulanarak özellik çıkarımı gerçekleştirilir. Özellik çıkarımı neticesinde ayırt ediciliği en yüksek olan özellikler (bantlar) belirlenmektedir. Veri setindeki değişkenlik bilgisini korumak için tüm bileşenler dikkate alınır. Özellik çıkarımı ile sınıflandırıcı kümesindeki her bir sınıflandırıcı için çeşitlilik korunmuş olur. RotOR algoritmasının temel çalışma prensibi aşağıda verilmiştir (Rodriguez et al., 2006):

X 'in eğitim veri setini, Y 'nin veri setindeki sınıf etiketlerini, F de özellik (bant) sayısını gösterdiği göz önüne alınsın. Eğitim veri setinin N sayıda pikselden oluştuğu ve n sayıda arazi örtüsü/kullanım sınıfı olduğu düşünülürse, eğitim veri seti (X), $N \times n$ boyutunda bir matris formundadır. Sınıf etiketlerini ifade eden Y vektörünün $[y_1, \dots, y_n]$ şeklinde olduğu düşünülür. RotOR algoritmasının çalışma prensibine uygun olarak, veri setinin yaklaşık olarak aynı boyutta K sayıda alt kümeye ayrıldığı, rotasyon ormanda D_1, \dots, D_L şeklinde gösterilen L sayıda karar ağacının olduğu göz önüne alınsın. Bu durumda rotasyon ormanındaki her bir karar ağacı D_i için eğitim veri seti sırasıyla aşağıdaki işlem adımları ile belirlenir:

1. Adım: F rastgele şekilde K sayıda bağımsız altkümeye bölünür. Her bir bağımsız alt küme $M = n/K$ sayıda özelliği (bandı) içermelidir.

2. Adım: F_{ij} 'nin D_i sınıflandırıcısının eğitiminde kullanılan j sayıda özelliği içeren alt kümesi ve X_{ij} 'nin X veri setinde F_{ij} deki özellikleri içeren veri seti olduğu düşünülün. Bu durumda yeni bir eğitim veri seti *bootstrap* yaklaşımı ile veri setinin %75 eğitim %25 test veri seti olarak belirlenir. Daha sonra yeni oluşturulan veri setine ana bileşen dönüşümü uygulanarak kovaryans matrisi C_{ij} hesaplanır.
3. Adım: Hesaplanan kovaryans değerleri kullanılarak Eşitlik 3'deki R_i dönüşüm matrisi oluşturulur.

$$R_i = \begin{bmatrix} a_{i1}^{(1)}, \dots, a_{i1}^{(M_1)} & \{0\} & \dots & \{0\} \\ \{0\} & a_{i2}^{(1)}, \dots, a_{i2}^{(M_2)} & \dots & \{0\} \\ \dots & \dots & \dots & \dots \\ \{0\} & \{0\} & \dots & a_{iK}^{(1)}, \dots, a_{iK}^{(M_K)} \end{bmatrix} \quad (3)$$

R_i matrisinin her bir sütunu orijinal özellik (bant) dizilimine göre yeniden düzenlenir ve yeniden düzenlenen rotasyon matrisi R_i^a elde edilir. Sonuç olarak D_i sınıflandırıcısının eğitiminde kullanılacak dönüştürülmüş veri seti XR_i^a olacaktır. Bu işlem adımları rotasyon ormanındaki tüm sınıflandırıcılar için tekrar edilir. Dönüştürülmüş veri setleri kullanılarak ormandaki her bir karar ağacına ait sınıflandırma sonucu bir oy alır. Sınıf etiketi bilinmeyen herhangi bir örnek (piksel), tüm ağaç tahminlerinde en fazla oy alan sınıfa atanması suretiyle sınıflandırılır.

4. UYGULAMA

Çalışma alanını kapsayan WorldView-2 görüntüsünün sınıflandırılması için öncelikli olarak 13 arazi örtüsü/kullanım sınıfını temsil eden örnekleme alanları belirlenmiştir. Söz konusu örnekleme alanları içerisinde rastgele örnekleme prensibinden hareketle her bir sınıf için eşit sayıda piksel içerecek şekilde eğitim verisi olarak toplam 6.500 piksel ve test veri seti olarak toplam 5.200 piksel belirlenmiştir. Rotasyon orman algoritmasının söz konusu veri setleri için sınıflandırma performansı hızlandırma ve rastgele orman algoritmalarının performansı ile karşılaştırılarak sonuçlar istatistiksel olarak analiz edilmiştir. Bunun yanında eğitim ve test veri setleri geleneksel sınıflandırıcı olarak bilinen en çok benzerlik algoritması ile de sınıflandırılarak topluluk öğrenme algoritmalarının sınıflandırma doğruluğu üzerindeki etkileri incelenmiştir.

En çok benzerlik (EÇB) yöntemiyle sınıflandırma işlemi MATLAB yazılımında hazırlanan bir program ile gerçekleştirilmiştir. Sınıflandırma sonucunda EÇB algoritmasıyla hesaplanan genel doğruluk Tablo 1'den de görüleceği üzere %83,42 ve Kappa istatistik değeri 0,82'dir.

Tablo 1. En çok benzerlik ve toplu öğrenme algoritmaları için hesaplanan sınıflandırma sonuçları.

Algoritma	Genel doğruluk (%)	Kappa değeri
En çok benzerlik	83,42	0,82
Hızlandırma	84,38	0,83
Rastgele orman	85,08	0,83
Rotasyon orman	86,17	0,85

Hızlandırma, rastgele orman ve rotasyon orman algoritmaları ile gerçekleştirilen sınıflandırma işleminde temel sınıflandırıcı olarak karar ağacı algoritmalarından C4.5 algoritması kullanılmıştır. Uygulamada dikkate alınan topluluk algoritmaları kullanıcı tarafından belirlenmesi gereken bir takım parametrelere sahiptir. Hızlandırma algoritması için ağırlıklı eşik değeri ve iterasyon sayısı en önemli iki parametre iken, rastgele orman algoritması için her bir düğümde kullanılan değişkenlerin sayısı ve ormandaki ağaç sayısı parametreleri mevcuttur. Rotasyon orman algoritması ile sınıflandırma işleminde ise alt küme sayısı ve ormandaki ağaç sayısı olarak adlandırılan iki parametre mevcuttur. Söz konusu parametre değerleri her bir algoritma için ayrı ayrı belirlenmiş ve hesaplanan sınıflandırma doğrulukları Tablo 1'de gösterilmiştir. Hızlandırma yönteminde ağırlıklı eşik değeri 100 ve optimum iterasyon sayısı 30 olarak tespit edilmiş ve bu değerler kullanılarak hesaplanan sınıflandırma doğruluğu tablodan da görüleceği üzere %84,38 ve Kappa değeri 0,83'dir. Rastgele orman (RO) algoritması her bir düğümdeki değişken sayısının 2 ve ormandaki ağaç sayısının 60 olarak seçildiği model ile en yüksek sınıflandırma doğruluğuna (%85,08)

ulaşmıştır. Rotasyon orman (RotOR) algoritmasının parametrelerinden alt küme sayısı 3 ve ormandaki ağaç sayısı 25 olarak belirlenmiş ve genel sınıflandırma doğruluğu %86,17 olarak hesaplanmıştır.

Çalışmada kullanılan en çok benzerlik (EÇB) ve toplu öğrenme algoritmalarının sınıflandırma performansları karşılaştırıldığında, değerlendirmeye alınan toplu öğrenme algoritmaları ile elde edilen sınıflandırma doğruluklarının EÇB ile elde edilen genel doğruluktan daha yüksek olduğu görülmektedir. Sınıflandırma doğrulukları arasındaki farklılıklar hızlandırma algoritması ile %0,96, RO algoritması ile %1,66 ve RotOR algoritması ile %2,75 olarak hesaplanmıştır. Doğrulukları arasındaki farklılıkların istatistiksel anlamlılığı McNemar testi kullanılarak analiz edildiğinde tüm farklılıkların %95 güven aralığında kritik tablo değeri 3,84'den büyük olduğu Tablo 2'de görülmektedir. Elde edilen bu sonuç sınıflandırma doğrulukları arasındaki farklılıkların istatistiksel olarak anlamlı olduğunu, dolayısıyla toplu öğrenme algoritmalarının sınıflandırma performansının EÇB'ye göre daha yüksek olduğunu desteklemektedir.

Tablo 2. En çok benzerlik ve toplu öğrenme algoritmalarının performans analizi için hesaplanan istatistik değerleri.

İstatistiksel Karşılaştırma	McNemar değeri
En çok benzerlik - Hızlandırma	3,95
En çok benzerlik - Rastgele orman	12,03
En çok benzerlik - Rotasyon orman	36,45

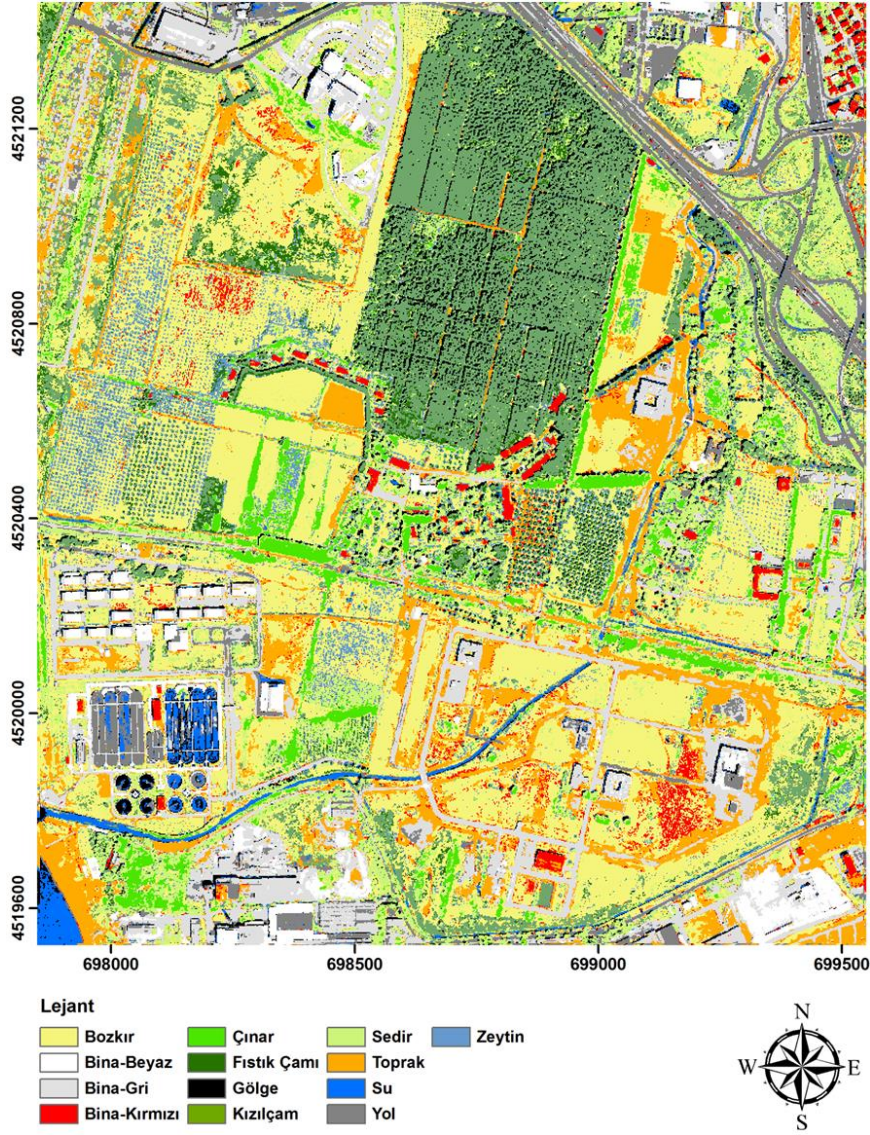
Çalışmada değerlendirmeye alınan toplu öğrenme algoritmalarının sınıflandırma performansları birbirleri ile karşılaştırıldığında en yüksek sınıflandırma doğruluğunun %86,17 ile RotOR algoritması ile elde edildiği görülmüştür. Yeni nesil toplu öğrenme algoritması olarak ifade edilen rotasyon orman algoritmasının sınıflandırma performansı hızlandırma ve rastgele orman algoritmalarının performansları ile karşılaştırıldığında, hızlandırma algoritmasından %1,79 rastgele orman algoritmasından %1,09 daha yüksek genel doğruluğa ulaşıldığı belirlenmiştir. Sınıflandırma doğrulukları arasındaki bu farklılıklar McNemar testi kullanılarak analiz edildiğinde %95 güven aralığında kiritik ki-kare tablo değerinden (3,84) büyük olduğu tespit edilmiş, dolayısıyla söz konusu farklılıkların istatistiksel olarak anlamlı olduğu sonucuna ulaşılmıştır (Tablo 3). Diğer bir ifadeyle uygulamada dikkate alınan veri seti için RotOR algoritması diğer toplu öğrenme algoritmalarından istatistiksel olarak daha yüksek sınıflandırma performansı sergilemiştir.

Tablo 3. Toplu öğrenme algoritmalarının performans analizi için hesaplan istatistik değerleri.

İstatistiksel Karşılaştırma	McNemar değeri
Rotasyon Orman - Hızlandırma	20,28
Rotasyon Orman - Rastgele orman	8,77
Hızlandırma - Rastgele orman	4,14

Toplu öğrenme algoritmaları içerisinde hesaplanan en düşük doğruluk %84,38 ile hızlandırma algoritması ile elde edilmiştir. Rastgele orman algoritmasının performansı ile karşılaştırıldığında genel doğruluklar arasında %0,70'lik bir fark olduğu görülmektedir. Söz konusu performans farklılığı Tablo 3'de verilen McNemar sonuçlarına göre istatistiksel olarak anlamlı bir fark olup, rastgele orman algoritmasının hızlandırma algoritmasından daha yüksek sınıflandırma performansı sergilediğini göstermektedir.

EÇB ve toplu öğrenme algoritmalarına ait elde edilen sınıflandırma modelleri kullanılarak WorldView-2 görüntüsü sınıflandırılmış ve çalışma alanına ait arazi örtüsü ve kullanımını gösteren tematik haritalar üretilmiştir. Söz konusu tematik haritalardan rotasyon orman algoritması ile üretileni Şekil 2'de gösterilmiştir. Şekil incelendiğinde özellikle toprak ve bina-kırmızı çatı olarak isimlendirilen sınıflara ait piksellerin çalışma alanının bazı bölgelerinde birbirlerine karıştığı görülmektedir. Bununla birlikte benzer spektral özelliklere sahip su ve gölge sınıflarına ait pikseller arasında da hatalı sınıflandırmalar olduğu görülmektedir. Ağaç türlerinin büyük oranda doğru sınıflandırıldığı, çalışma alanının belirli kesimlerinde canlı bitki örtüsünü içerisinde barındıran bozkır sınıfına ait piksellerin geniş yapraklı ağaç türlerini temsil eden arazi örtüsü sınıfları içerisine dahil edildiği tespit edilmiştir.



Şekil 2. WorldView-2 görüntüsünün RotOR algoritması ile sınıflandırılması sonucu elde edilen tematik harita.

5. SONUÇLAR

Uzaktan algılama teknolojilerinde yaşanan yenilikler kullanıcıları yeni nesil uydu görüntülerinin analizi noktasında güçlü ve etkili alternatif yöntemlerin geliştirilmesi arayışına itmiştir. Özellikle uydu görüntülerinin sınıflandırılması ve tematik harita üretiminde son yıllarda istatistiksel algoritmalar yerine parametrik olmayan, karmaşık yapıdaki piksellerin sınıflandırılmasında hızlı ve etkili sınıflandırma algoritmaları geliştirilmiştir. Bunlar arasında toplu öğrenme algoritmaları başta uzaktan algılama olmak üzere birçok alanda başarıyla kullanılmaktadır. Son yıllarda rotasyon algoritması (RotOR) olarak isimlendirilen toplu öğrenme algoritması uzaktan algılanmış görüntülerin sınıflandırılması problemini de içerisine alan birçok uygulama probleminin çözümünde kullanılan etkili bir sınıflandırma algoritması olarak önerilmektedir.

Bu çalışmada RotOR algoritmasının yüksek çözünürlüklü WV-2 uydu görüntüsünün sınıflandırılmasındaki performansı incelenmiş ve elde edilen sonuçlar popüler toplu öğrenme algoritmalarından hızlandırma ve rastgele orman algoritmaları ile karşılaştırılarak analiz edilmiştir. RotOR ve diğer toplu öğrenme algoritmalarının istatistiksel bir metot olan en çok benzerlik (EÇB) yöntemi ile karşılaştırılarak algoritmaların geleneksel yöntemlere göre performansları da analiz edilmiştir. Çalışma sonucunda yeni nesil toplu öğrenme algoritması olarak önerilen RotOR algoritması kullanılarak popüler toplu öğrenme algoritmalarına göre %2'lere varan seviyelerde daha doğru sınıflandırma sonuçlarına ulaşıldığı görülmüştür. Elde edilen bu sonuç McNemar testi kullanılarak analiz edilmiş ve RotOR algoritması ile sınıflandırma performansında istatistiksel olarak anlamlı artışlar olduğu tespit edilmiştir. Diğer taraftan hızlandırma ve rastgele orman algoritmaları için elde edilen sınıflandırma doğrulukları arasındaki farklılık istatistiksel olarak anlamlı bulunmuş ve uygulamada dikkate alınan veri seti için rastgele orman algoritmasının hızlandırma algoritmasından daha yüksek sınıflandırma doğruluğu ürettiğini desteklemektedir.

Çalışmada değerlendirmeye alınan toplu öğrenme algoritmalarının sınıflandırma performansı istatistiksel kabullere dayanan EÇB algoritması ile karşılaştırıldığında, toplu öğrenme yaklaşımıyla %3'e varan ve istatistiksel olarak anlamlı performans farklılıkları elde edildiği tespit edilmiştir. Yapılan performans analizleri ve elde edilen istatistiksel test sonuçları rotasyon orman algoritmasının yüksek çözünürlüklü uydu görüntülerinin sınıflandırılması ve karmaşık yapıdaki arazi örtüsü/kullanımına ilişkin tematik harita üretiminde kullanılabilecek etkili ve alternatif bir yöntem olduğunu destekler niteliktedir.

KAYNAKLAR

- Breiman, L.**, 2001, Random Forests, *Machine Learning*, 45(1), 5-32.
- Çölkesen, İ., Kavzoğlu, T.**, 2009, İyileştirilmiş Karar Ağaçları ile Uydu Görüntülerinin Sınıflandırılması, *TMMOB Harita ve Kadastro Mühendisleri Odası 12. Türkiye Harita Bilimsel ve Teknik Kurultayı*, 11-15 Mayıs 2009, Ankara.
- Demarchi, L., Canters, F., Cariou, C., Licciardi, G., Chan, J.C.W.**, 2014, Assessing The Performance of Two Unsupervised Dimensionality Reduction Techniques on Hyperspectral APEX Data For High Resolution Urban Land-Cover Mapping, *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 166-179.
- Dietterich, T.G.**, 2000a, Ensemble Methods in Machine Learning. In J. Kittler and F. Roli (Ed.) *First International Workshop on Multiple Classifier Systems*, Lecture Notes in Computer Science (pp. 1-15). New York: Springer Verlag.
- Dietterich, T.G.**, 2000b, An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization, *Machine Learning*, 40(2), 139-158.
- Foody, G.M.**, 2004, Thematic Map Comparison: Evaluating The Statistical Significance of Differences in Classification Accuracy, *Photogrammetric Engineering and Remote Sensing*, 70(5), 627-633.
- Freund, Y., Schapire, R.E.**, 1996, Experiments with A New Boosting Algorithm, In: Saitta, L. (Ed.), *Proceedings of the Thirteenth International Conference on Machine Learning (ICML-96)*, Morgan Kaufmann, pp.148-156.
- Garrity, S.R., Allen, C.D., Brumby, S.P., Gangodagamage, C., McDowell, N.G., Cai, D.M.**, 2013, Quantifying Tree Mortality in A Mixed Species Woodland Using Multitemporal High Spatial Resolution Satellite Imagery, *Remote Sensing of Environment*, 129, 54-65.
- Ghosh, A., 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.
- Hamedianfar, A., Shafri, H.Z.M., Mansor, S., Ahmad, N.**, 2014, Improving Detailed Rule-Based Feature Extraction of Urban Areas from Worldview-2 Image and Lidar Data. *International Journal Of Remote Sensing*, 35(5), 1876-1899.
- Hayes, M.M., Miller, S.N., Murphy, M.A.**, 2014, High- resolution Land Cover Classification Using Random Forest. *Remote Sensing Letters*, 5, 112-121.
- Huang, X., Lu, Q.K., Zhang, L.P.**, 2014, A Multi-Index Learning Approach for Classification of High-Resolution Remotely Sensed Images Over Urban Areas, *ISPRS Journal of Photogrammetry and Remote Sensing*, 90, 36-48.
- Kavzoğlu, T., 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.
- 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.
- Lu, D.S., Hetrick, S., Moran, E.**, 2010, Land Cover Classification in A Complex Urban-Rural Landscape with Quickbird Imagery, *Photogrammetric Engineering and Remote Sensing*, 76(10), 1159-1168.

- Malinverni, E. S., Tassetti, A. N., Mancini, A., Zingaretti, P., Frontoni, E., Bernardini, A.,** 2011, Hybrid Object-Based Approach for Land Use/Land Cover Mapping Using High Spatial Resolution Imagery, *International Journal of Geographical Information Science*, 25(6), 1025-1043.
- Mather, P.M.,** 1999, *Computer Processing of Remotely Sensed Images*, Second Edition (Chichester: John Wiley & Sons).
- Mei, A., Salvatori, R., Fiore, N., Allegrini, A., D'Andrea, A.,** 2014, Integration of Field and Laboratory Spectral Data with Multi-Resolution Remote Sensed Imagery for Asphalt Surface Differentiation. *Remote Sensing*, 6(4), 2765-2781.
- Nanni, L., Lumini A.,** 2006, An Experimental Comparison of Ensemble of Classifiers for Biometric Data, *Neurocomputing*, 69(13-15), 1670-1673.
- Opitz, D., Maclin R.,** 1999, Popular Ensemble Methods: An Empirical Study, *Journal of Artificial Intelligence Research*, 11, pp. 169–198.
- 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), 554-565.
- Patino, J.E., Duque, J.C.,** 2013, A Review of Regional Science Applications of Satellite Remote Sensing in Urban Settings. *Computers Environment and Urban Systems*, 37, 1-17.
- Rodriguez, J.J., Kuncheva, L.I. and Carlos, J.A.,** 2006, Rotation Forest: A New Classifier Ensemble Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10), 1619-1630.
- Schapire, R.E.,** 1999, Theoretical Views of Boosting And Applications, In: *Proceedings of the Tenth International Conference on Algorithmic Learning Theory*, pp. 13-25.
- Singh, M., Malhi, Y., Bhagwat, S.,** 2014, Biomass Estimation of Mixed Forest Landscape Using A Fourier Transform Texture-Based Approach on Very-High-Resolution Optical Satellite Imagery. *International Journal of Remote Sensing*, 35(9), 3331-3349.
- Waser, L.T., Kuchler, M., Jutte, K., Stampfer, T.,** 2014, Evaluating The Potential of Worldview-2 Data to Classify Tree Species and Different Levels of ASH Mortality, *Remote Sensing*, 6(5), 4515-4545.