

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/283509319>

UZAKTAN ALGILANMIŞ GÖRÜNTÜLERDE OPTİMUM BANTLARIN SEÇİMİNDE DESTEK VEKTÖR MAKİNELERİNİN KULLANIMI (The use of support vector machine for optimal band selection for remotely sensed...

Conference Paper · May 2015

CITATIONS

0

READS

44

3 authors:



Ismail Colkesen

Gebze Technical University

31 PUBLICATIONS 325 CITATIONS

SEE PROFILE



Taskin Kavzoglu

Gebze Technical University

62 PUBLICATIONS 1,076 CITATIONS

SEE PROFILE



Tahsin Yomralioglu

Istanbul Technical University

53 PUBLICATIONS 307 CITATIONS

SEE PROFILE

UZAKTAN ALGILANMIŞ GÖRÜNTÜLERDE OPTİMUM BANTLARIN SEÇİMİNDE DESTEK VEKTÖR MAKİNELERİNİN KULLANIMI

İ. Çölkesen^{a,*}, T. Kavzoğlu^a, T. Yomralıoğlu^b

^aGTÜ, Jeodezi ve Fotogrametri Müh. Böl., 41400 Gebze-Kocaeli, Türkiye - (icolkesen, kavzoglu)@gtu.edu.tr

^bİTÜ, Geomatik Müh. Böl., 34469 Maslak, İstanbul, Türkiye – tahsin@itu.edu.tr

ANAHTAR SÖZCÜKLER: Sınıflandırma, WorldView-2, Vejetasyon İndeksleri, Özellik Seçimi, Destek Vektör Makineleri

ÖZET:

Yüksek konumsal çözünürlüğe sahip uydu görüntüleri yardımıyla yeryüzü nesnelerin konumlarına ve dağılımlarına ilişkin detaylı bilgilerin elde edilmesi mümkündür. Buna karşın yüksek çözünürlüklü görüntüler heterojen yapıda ve benzer yansıma değerlerine sahip yoğun görüntü piksellerinden oluşmaktadır. Karmaşık yapıdaki pikseller arasındaki spektral ayrımın artırılması amacıyla uydu görüntüleri ile birlikte yardımcı verilerin bir arada kullanılması gereklidir. Entegrasyon sonucunda elde edilen veri seti yüksek boyutlu ve korelasyonlu bantları (özellikleri) içermektedir. Bu nedenle yüksek boyutlu veri seti içerisinde sınıflandırma probleminin çözümüne katkı sağlayacak bantların belirlenmesi önemli bir araştırma konusudur. Bu çalışmada, yüksek çözünürlüğe sahip WorldView-2 (WV-2) uydu görüntüsü ile birlikte 12 farklı vejetasyon indeksi, temel bileşenler analizi sonucunda hesaplanan ilk üç bileşen, gri-eş-düzye oluşum matrisi yardımıyla hesaplanan doku özellikleri ve ton-doygunluk-yoğunluk dönüşümü ile hesaplanan bileşenler yardımcı veri seti olarak değerlendirilmeye alınmıştır. Yüksek boyutlu veri seti içerisinde optimum bantların seçiminde özyinelemeli özellik seçimi esasına dayanan destek vektör makineleri (DVM-ÖÖE) kullanılmıştır. Özellik seçimi sonucunda elde edilen veri setlerinin sınıflandırılmasında geleneksel en yakın komşuluk (EYK) algoritması ve ileri sınıflandırma algoritmalarından destek vektör makineleri (DVM) kullanılmıştır. Sonuçlar, DVM algoritması ile hesaplanan en yüksek genel sınıflandırma doğruluğunun (%88,23) seçilen 40 bantı içeren veri seti ile elde edildiğini göstermektedir. Diğer taraftan, EYK algoritması en yüksek sınıflandırma doğruluğuna (%83,74) seçilen 20 bantla ulaşmıştır. McNemar's testi sonuçlarına göre sınıflandırma doğrulukları arasındaki farkın istatistiksel olarak anlamlı olduğu tespit edilmiştir. Çalışma sonuçları bir bütün olarak ele alındığında, DVM-ÖÖE algoritmasıyla veri seti boyutu %60-80 azaltılarak sınıflandırma doğruluğunda %5'e varan seviyelerde iyileşme olabileceğini göstermektedir.

KEY WORDS: Classification, WorldView-2, Vegetation Indices, Feature selection, Support Vector Machines

ABSTRACT:

High resolution satellite images provide more detailed information about the location and distribution of the Earth's surface objects. However, these images consist of intensive heterogeneous image pixels having similar reflectance values. In order to increase the spectral separability between the pixels with complex structure, auxiliary data sets are used in combination with satellite images. Data set resulting from this integration comprises high dimensional and correlated image bands (features). For this reason, determination of optimum bands contributing the solution of the classification problems is an important research topic. In this study, in addition to WorldView-2 (WV-2) image bands, 12 vegetation indices, first three principal components, texture features calculated by using gray-level co-occurrence matrix and components of hue-saturation-intensity transformation were considered as an auxiliary data. Support vector machines based on recursive feature elimination (SVM-RFE) was used for selecting optimum bands within the high dimensional data. Traditional nearest neighbor (NN) and support vector machines (SVMs) classifiers were applied for the classification of data sets obtained after the feature selection process. Results showed that the SVMs algorithm produced the highest classification accuracy of 88.23% with the data set including 40 selected bands. On the other hand, NN algorithm produced the best classification performances (83.74%) with 20 selected bands. According to the McNemar's test results, the difference in classification accuracies was statistically significant. All in all, with the use of SVM-RFE algorithm the improvement in classification accuracy reached to 5% by reducing the data set size by about 60-80%.

1. GİRİŞ

Uzaktan algılama teknolojileri yeryüzü nesnelerinin niteliği, çeşitliliği ve konumsal dağılımlarına ilişkin önemli bilgilerin elde edilmesinde etkili ve ekonomik çözümler sunmaktadır. Son yıllarda algılayıcı sistemlerinin tasarımında yaşanan yenilikler yeryüzünün yüksek konumsal çözünürlüklerde görüntülenmesine imkan sağlamıştır. Yeni nesil uydu görüntüleri yeryüzü nesnelere ilişkin detaylı bilgiler içermesine karşın, sağladığı spektral bilgi içeriği sınırlı seviyededir (Carleer et al., 2005). Bu nedenle, yüksek

çözünürlüklü görüntüler üzerinde karmaşık yapıda ve benzer spektral özelliklere sahip pikseller arasındaki spektral ayrımın artırılması amacıyla yardımcı veri seti kullanımı söz konusu olmuştur.

Multispektral görüntüler üzerindeki bilgi içeriğinin bir kısmının veya tamamının yeniden düzenlenerek daha anlamlı hale getirilmesi olarak tanımlanan görüntü dönüşümleri yardımcı veri seti oluşturulmasında en çok tercih edilen yöntemdir (Mather and Koch, 2011; Richards, 2013). Multispektral görüntünün bir bandını veya birden fazla bandını kullanarak

* Sorumlu yazar. Tel: 0.262.605 1808

dönüşüm işleminin gerçekleştirilmesinde toplama, çıkarma, çarpma ve bölme işlemlerini içeren aritmetik operasyonlar ve bunlar dışında kalan birçok hesaplama ölçütü kullanabilmektedir. Literatürde vejetasyon indekslerinin, doku özelliklerinin, temel bileşenler analizlerinin ve ton-doygunluk-yoğunluk dönüşümlerinin en çok kullanılan görüntü dönüşüm teknikleridir (Lu et al., 2010, Low et al., 2013, Shamsoddini et al., 2013; Ghosh and Joshi, 2014; Çölkesen ve Yomralıoğlu, 2014).

Uydu görüntüleri ve yardımcı verilerin bir arada kullanılması durumunda yüksek boyutlu veri setinin sınıflandırılması problemi ortaya çıkmaktadır. Bunun temel nedeni yüksek boyutlu veri seti içerisinde sınıflandırma probleminin çözümüne fayda sağlamayan ya da birbirleriyle yüksek kolerasyona sahip bantların mevcut olmasıdır. Bununla birlikte yüksek boyutlu verilerle yapılan sınıflandırma işleminde boyutun artmasına paralel olarak belirli bir noktadan sonra sınıflandırma doğruluğunun düştüğü bilinmektedir. Yüksek boyutlu verilerde ortaya çıkan bu durum Hughes fenomeni veya boyutsallık kavramları ile açıklanmaktadır (Hughes, 1968; Mianji ve Zhang, 2011). Bu nedenle yüksek boyutlu verilerin değerlendirilmesi aşamasında hem işlem yükünün azaltılması hem de boyutsallık probleminin önüne geçilerek sınıflandırma doğruluğunun artırılması amacıyla özellik seçimi yoluna gidilmektedir. Literatürde özellik seçimine yönelik birçok algoritmanın geliştirildiği ve farklı uygulama alanlarında kullanıldığı görülmektedir (Kavzoglu and Mather, 2000; Baccianella et al., 2014; Fassnacht et al., 2014; Mwangi et al., 2014). Birçok sınıflandırma ve örüntü tanıma probleminin çözümünde başarıyla kullanılan destek vektör makineleri son yıllarda özellik seçimi ve boyut indirgeme problemlerinin çözümünde de kullanılmaya başlanmıştır (Zhang and Ma, 2009; Pal and Foody, 2010; Atkinson et al., 2014).

Bu çalışmada yüksek çözünürlüklü uydu görüntüsü üzerinden arazi örtüsü ve arazi kullanımına ilişkin tematik harita üretiminde temel veri seti olarak 2013 yılında kaydedilen WorldView-2 (WV-2) uydu görüntüsü kullanılmıştır. Uydu görüntüsünün sahip olduğu 8 spektral banda ilave olarak 12 farklı vejetasyon indeksi, temel bileşenler analizi sonucunda hesaplanan ilk üç bileşen, gri eş düzey oluşum matrisi yardımıyla hesaplanan doku özellikleri ve ton-doygunluk-yoğunluk dönüşümü ile hesaplanan bileşenler yardımcı veri seti olarak değerlendirmeye alınmıştır. WV-2 ve yardımcı verilerin entegrasyonu neticesinde 96 bantlı (özellığı) içeren yüksek boyutlu veri seti elde edilmiştir. Söz konusu veri seti içerisinde sınıflandırma doğruluğuna katkı sağlayan optimum bant sayısının tespit edilmesi amacıyla özyinelemeli özellik seçimi esasına dayanan destek vektör makineleri kullanılmıştır. Bu amaca yönelik olarak özellik seçimi sonucunda belirlenen önem sıralamaları dikkate alınarak farklı boyutlarda veri setleri oluşturulmuştur. Söz konusu veri setlerinin sınıflandırılmasında en yakın komşuluk (EYK) algoritmasının yanında makine öğrenme algoritması olarak bilinen destek vektör makineleri (DVM) kullanılmıştır. Çalışma sonuçlarının analizinde, genel sınıflandırma doğruluklarının karşılaştırılmasının yanı sıra doğruluklar arasındaki farklar McNemar's testi kullanılarak istatistiksel olarak analiz edilmiştir.

2. ÇALIŞMA ALANI VE KULLANILAN VERİ

Kocaeli iline bağlı Gebze ilçesinin batı kısmında, İstanbul ili sınırında yer alan ve Gebze Teknik Üniversitesi (GTÜ) Çayırova kampüsünü de kapsayan yaklaşık 220 hektarlık 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.



Şekil 1. Çalışma alanına ait WV-2 görüntüsü.

Uygulamaya esas veri seti olarak 2013 tarihinde kaydedilen WV-2 uydu görüntüsü kullanılmıştır. WV-2 uydusu 2m konumsal çözünürlüğe sahip, kıyı (0.40–0.45 μm), mavi (0.45–0.51), yeşil (0.51–0.58 μm), sarı (0.59–0.63 μm), kırmızı (0.63–0.69 μm), yakın kızılötesi-1 (0.77–0.90 μm) ve yakın kızılötesi-2 (0.86–1.04 μm) olarak bilinen 8 spektral bantta görüntüleme yapabilmektedir. Bununla birlikte WV-2 0,5 m konumsal çözünürlüğe sahip pankromatik banda (0.46–0.80 μm) sahiptir. Uydu görüntüsünün geometrik düzeltilmesi ve UTM koordinat sistemine dönüşümünde mevcut hâlihazır haritalar ve bölgeye ait yüksek çözünürlüklü görüntülerden yararlanılmıştır. WV-2 görüntüsünün pankromatik bandı ile multispektral bantlarının kaynaştırılmasında Gram-Schmidt algoritması kullanılmıştır. Kaynaştırma sırasında 3x3 boyutlu yumuşatma filtresi uygulanmış ve yeniden örnekleme işleminde bilineer enterpolasyon algoritması kullanılmıştır.

3. YARDIMCI VERİLERİN OLUŞTURULMASI

Çalışma amacına uygun olarak WV-2'nin mevcut spektral bantlarına ilave olarak vejetasyon indeksleri, temel bileşenler analizi, doku özellikleri ve ton-doygunluk-yoğunluk dönüşümü neticesinde elde edilen bileşenler yardımcı veri seti olarak değerlendirilmeye alınmıştır.

Bitki canlılığının belirlenmesinde kullanılan niceliksel ölçümler olarak tanımlanan vejetasyon indeksleri, farklı spektral bantların sahip olduğu yansıma değerlerinin doğrusal kombinasyonları veya oranlarını içerisine alan matematiksel dönüşümler olarak bilinmektedir (Campbell and Wynne, 2011). Günümüze kadar literatürde birçok vejetasyon indeksi tanımlanmış ve uygulamalarda kullanılmıştır (Bannari et al., 1995; Payero et al., 2004; Waser et al., 2014). Bu çalışmada, yardımcı veri seti olarak WV-2 görüntüsünün yeni bantlarının da kullanıldığı 12 farklı vejetasyon indeksi (Tablo 1) hesaplanmış ve her bir indeks için tematik harita üretilmiştir.

Tablo 1. Vejetasyon indeksleri ve formülleri.

Vejetasyon İndeksi	Formül
Normalleştirilmiş fark bitki örtüsü indeksi - 1	$\frac{NIR2 - Kırmızı}{NIR2 + Kırmızı}$
Normalleştirilmiş fark bitki örtüsü indeksi - 2	$\frac{NIR1 - Kırmızı}{NIR1 + Kırmızı}$
Normalleştirilmiş fark bitki örtüsü indeksi - 3	$\frac{NIR2 - Kırmızı kenar}{NIR2 + Kırmızı kenar}$
Normalleştirilmiş toprak indeksi	$\frac{Sarı - Yeşil}{Sarı + Yeşil}$
Homojen olmayan özellik farkı	$\frac{Kırmızı kenar - Kıyı}{Kırmızı kenar + Kıyı}$
Geliştirilmiş bitki örtüsü indeksi	$2,5 \times \frac{NIR2 - Kırmızı}{NIR2 + 6 \times NIR2 - 7,5 \times Mavi + 1}$
Düzenlenmiş toprak etkisi azaltılmış vejetasyon indeksi	$\frac{2 \times NIR2 + 1 - \sqrt{(2 \times NIR2 + 1)^2 - 8 \times (NIR2 - Kırmızı)}}{2}$
Orman ve bitki indeksi	$\frac{NIR1 - Kırmızı kenar}{NIR1 + Kırmızı kenar}$
Toprak etkisi azaltılmış vejetasyon indeksi	$(1 + L) \times \frac{NIR2 - Kırmızı}{NIR2 + Kırmızı + L}$
Atmosferik koşullara dayanıklı bitki indeksi	$\frac{NIR2 - 2 \times Kırmızı + Mavi}{NIR2 + 2 \times Kırmızı - Mavi}$
Optimize edilmiş toprak etkisi azaltılmış vejetasyon indeksi	$\frac{NIR2 - Kırmızı}{NIR2 + Kırmızı + 0,16}$
Temel bileşen ve kızılötesi bant ile normalleştirilmiş indeks	$\frac{PCAI - NIR2}{PCAI + NIR2}$

*NIR1: Yakınkızılötesi-1, NIR2: Yakınkızılötesi-2, PCA1: İlk temel bileşen, L: Düzeltme faktörü

Temel bileşenler analizi (TBA) genel olarak bantlar arasındaki korelasyonun yok edilmesi, boyut indirgeme veya azaltma amacıyla yaygın olarak kullanılan bir analiz tekniğidir (Lillesand et al., 2007). TBA ile korelasyonlu birçok spektral bantı temsil eden daha az sayıda korelasyonsuz bileşenler tanımlanmaktadır. Bu işlemin gerçekleştirilmesinde söz konusu veri setinin ortogonal düzleme matematiksel izdüşümü gerçekleştirilir. Bu açıdan ele alındığında temel bileşenler analizinde amaç değişimi ifade edecek yeni bir koordinat sisteminin tanımlanmasıdır (Rencz and Ryerson, 1999). TBA sonucunda bileşenlere ait hesaplanan varyans değerleri söz konusu bileşenin bilgi içeriğini gösterirken, genellikle ilk üç bileşen toplam bilginin yaklaşık %90'dan fazlasını içermektedir. WV-2 görüntüsü kullanılarak gerçekleştirilen temel bileşenler analizi sonucunda ilk üç bileşenin toplamda %98,84 varyansa sahip olduğu görülmüş ve uygulamada söz konusu bileşenler dikkate alınmıştır.

Düzenli bir frekansta tekrar eden desen ve/veya desen kombinasyonu olarak tanımlanan doku, bir görüntüdeki ilgi duyulan objeleri tanımlanmasında kullanılan önemli bir

özelliştir. Doku özellikleri iki komşu piksel arasındaki uzaklık ve açı ilişkilerinin bir fonksiyonu olarak tanımlanan gri düzey eş oluşum matrislerinden hesaplanmaktadır (Haralick et al., 1973). Gri düzey eş oluşum matrisleri belirli bir mesafede ve yönde bulunan piksel çiftleri arasında gri seviyesinin i 'den j 'ye geçen olasılıklarını gösteren iki boyutlu bir matristir. Matrisin hesaplanmasında pikseller arasındaki gri seviyesindeki değişimlerin belirlenmesinde mesafe olarak genellikle 3x3 veya 5x5 gibi pencere boyutu kullanılırken, desen özelliklerinin yönelimine göre genellikle görüntü üzerinde dört yönde (0°, 45°, 90° ve 135°) inceleme yapılır. Gri düzey eş oluşum matrisi yardımıyla hesaplanabilen doku özellikleri ortalama, varyans, homojenlik, zıtlık, farklılık, entropi, ikinci moment ve korelasyon olarak adlandırılmaktadır. Çalışma kapsamında, WV-2 uydusu için söz konusu 8 doku özelliğinin hesaplanması için 3x3 pencere boyutu dört farklı açıda (0°, 45°, 90° ve 135°) görüntüye uygulanmıştır.

Renk uzayları renkleri tanımlamak için kullanılan matematiksel modeller olup, bütün renkleri temsil edecek şekilde 3 boyutlu olarak tanımlanmaktadır. Bunun temel nedeni herhangi bir rengin elde edilmesinde üç değişkene ihtiyaç duyulmasıdır. Renklerin renk uzayındaki konumları söz konusu üç değişkenin alacağı değerlere göre belirlenir. Uygulamalarda en çok bilinen ve kullanılan renk uzayı kırmızı, yeşil ve mavi ana renklerinden oluşan RGB uzayıdır. Uzaktan algılama uygulamalarında uzaktan algılanmış görüntülerin analizine fayda sağlaması amacıyla RGB renk uzayından başka bir renk uzayına dönüşüm yapılabilmektedir. Uydu görüntüsünün yorumlanabilirliğini arttırmak amacıyla dönüşüm yapılan renk uzaylarının başında ton-doygunluk-yoğunluk (Hue-Saturation-Intensity, HSI) uzayı gelmektedir (Mather and Koch, 2011). HSI uzayında ton görünen veya algılanan rengin baskın dalga boyunu, doymunluk bir rengin saflık derecesini veya renk içerisindeki beyaz miktarını, yoğunluk ise bir rengin parlaklığını ya da matlığını göstermektedir. RGB uzayından HSI uzayına dönüşüm neticesinde RGB küpündeki 0-255 aralığındaki değerler, hexacone uzay modelinde doymunluk ve yoğunluk için 0-1 aralığında değerler alırken, ton için 0° - 360° açısal değerler almaktadır. Çalışma kapsamında, ton-doygunluk-yoğunluk dönüşümünde WV-2 görüntüsünün 3 farklı bant kombinasyonu kullanılmıştır. Bunlardan ilki WV-2 görüntüsünün klasik bantlarından (kırmızı (R)-yeşil (G)-mavi (B)) oluşurken, diğer ikisi WV-2'nin kırmızı ve yeşil bantlarıyla birlikte kızılötesi-1 (NIR1) ve kızılötesi-2 (NIR2) bantlarının kullanıldığı kombinasyonlardır (NIR1-R-G ve NIR2-R-G).

4. ÖZELLİK SEÇİMİ VE BOYUT İNDİRGEME

WV-2 uydu görüntüsünün sahip olduğu 8 spektral bant ve yardımcı veriler entegre edilerek yüksek boyuta (bant sayısına) sahip veri seti elde edilmiştir. Veri setinin sınıflandırması sürecinde hem işlem yükünün azaltılması hem de boyutsallık probleminin önüne geçilerek sınıflandırma doğruluğunun artırılması amacıyla boyut indirgemesi yoluna gidilmiştir. Boyut azaltılması ya da indirgemesi bant seçimi veya özellik çıkarımı olarak bilinen yaklaşımlarla gerçekleştirilmektedir. Boyut azaltılması neticesinde yüksek boyutlu veri setini temsil eden daha düşük boyutlu bir veri seti tanımlanmaktadır. Söz konusu veri seti düşük korelasyona sahip ve arazi örtüsü/kullanım sınıfları arasındaki spektral ayırımı optimum şekilde yapılabilmesine olanak sağlayacak yararlı spektral bilgileri içeren seçilmiş bantlardan oluşmaktadır.

Özyinelemeli özellik eleme esasına dayalı destek vektör makineleri (DVM-ÖÖE) algoritması, destek vektör makinelerini temel sınıflandırıcı olarak kullanan sarmal tabanlı özellik seçimi algoritmasıdır (Guyon et al., 2002; Pal and Foody, 2010). Algoritma veri seti içerisindeki özelliklerin sınıfları ayırt edebilme kabiliyetine göre sıralanmasında DVM sınıflandırıcısının temel amaç fonksiyonu olan $(1/2)\|w\|^2$ eşitliğini özellik sıralama kriteri olarak kullanmaktadır. Her bir işlem adımında, ağırlık vektörüne (w) ait katsayılar tüm özelliklerin sıralama puanlarının hesaplanmasında kullanılır. Bu açıdan ele alındığında, en küçük sıralama puanına $((w_i)^2)$ sahip olan özellik elenir. Burada w_i , i . özelliğin ağırlığını göstermektedir. Bu özellik seçimi yaklaşımı bir veri seti içerisindeki özelliklerin önem sırasına göre sıralandırılması işlemlerinin gerçekleştirilmesinde özyinelemeli yaklaşımla anlamsız özelliklerin belirlenmesinde geri yönlü eliminasyon semasını kullanmaktadır.

5. SINIFLANDIRMA ALGORİTMALARI

Bu çalışmada WorldView-2 ve yardımcı veri setlerinin entegrasyonu neticesinde elde edilen yüksek boyutlu veri setinin ve özellik seçimi sonucunda üretilen farklı boyutlarda veri setlerinin sınıflandırılmasında destek vektör makineleri ve en yakın komşuluk algoritmaları kullanılmıştır.

5.1 Destek Vektör Makineleri

Destek vektör makineleri (DVM) uzaktan algılanmış görüntülerin sınıflandırılması amacıyla son yıllarda yaygın bir şekilde kullanılan ve sınıflandırmadaki etkinliği birçok çalışmada ortaya konulan bir sınıflandırma algoritmasıdır (Huang et al., 2002; Kavzoglu and Colkesen, 2009; Mountrakis et al., 2011). DVM parametrik olmayan ve istatistiksel öğrenme teorisine dayanan yeni nesil bir sınıflandırma algoritmasıdır. Algoritmanın temel çalışma prensibi iki sınıfa ait pikselleri birbirinden optimum şekilde ayırabilen bir hiperdüzlemin belirlenmesi esasına dayanmaktadır (Vapnik, 1995).

Destek vektör makineleri ile uzaktan algılama uygulamalarında olduğu gibi doğrusal olarak ayrılamayan verilerin sınıflandırılması probleminin çözümünde kernel fonksiyonlarından yararlanır. Kernel fonksiyonları veri setini daha yüksek boyutlu bir uzayda görüntüleyerek bu uzayda verilerin doğrusal olarak ayrılabilmesine imkan sağlamaktadır (Kavzoglu and Colkesen, 2009). Bu çalışmada kernel fonksiyonu olarak sınıflandırma problemindeki etkinliği nedeniyle radyal tabanlı fonksiyon tercih edilmiştir.

5.2 En Yakın Komşuluk Algoritması

En temel örnek tabanlı öğrenme algoritmalarından birisi olan en yakın komşuluk (EYK) algoritması, uzaktan algılama alanında geleneksel sınıflandırıcı olarak kabul edilen parametrik bir algoritmadır. EYK algoritması öncelikli olarak eğitim kümesi içerisindeki sınıfları temsil eden ortalama değer vektörleri hesaplar. Daha sonra sınıflandırılmak istenen aday pikselin hesaplanan sınıf ortalama vektörleri arasındaki spektral uzaklıklar hesaplanır. Hesaplanan spektral uzaklıklara göre aday piksel en yakın mesafedeki örnek sınıfa atanır (Lillesand et al., 2007).

6. UYGULAMA

Sınıflandırmaya esas olacak veri seti Tablo 2'den de görüleceği üzere, WV-2 uydu görüntüsünün 8 spektral bandını, yardımcı veri olarak tanımlanan 12 vejetasyon indeksini, temel bileşenler analizi ile hesaplanan ilk üç bileşeni, uydu görüntüsünün spektral bantları kullanılarak hesaplanan 8 doku özelliğini ve WV-2'nin kırmızı, yeşil, mavi, kızılötesi-1 ve kızılötesi-2 bantlarının çeşitli kombinasyonları kullanılarak oluşturulan 3 farklı ton-doğunluk-yoğunluk özelliğinden oluşan 96 özelliği (bandı) içermektedir.

Sınıflandırma işleminin gerçekleştirilmesi ve özellik seçiminde kullanılmak üzere çalışma alanı içerisinde tespit edilen 13 farklı arazi örtüsü/kullanım sınıfını temsil eden örnekleme alanları tespit edilmiştir. Geniş ve iğne yapraklı sınıflar için örnekleme alanları arazide ASD FieldSpec-3 spektrometre cihazı ile gerçekleştirilen spektral ölçüler yardımıyla tespit edilirken, diğer sınıflar için örnekleme alanları mevcut hava fotoğrafları, uydu görüntüleri ve tematik haritalar yardımıyla tespit edilmiştir. Belirlenen örnekleme alanları içerisinde MATLAB yazılımında hazırlanan bir program ile rastgele örnekleme prensibinden hareketle 13 arazi örtüsü/kullanımı sınıfı için eşit sayıda örnek pikseller seçilerek eğitim ve test veri setleri oluşturulmuştur. Sonuç olarak eğitim verisi olarak her bir sınıf için 200 piksel içerecek şekilde toplam 2.600 piksel, test verisi olarak her bir sınıf için 150 pikseli içerecek şekilde toplam 1.950 piksel belirlenmiştir.

Tablo 2. Uygulamada kullanılan veri seti.

Özellik (bant)	Açıklama
Spektral bantlar	WV-2 bantları (8 bant)
Vejetasyon indeksleri	Vejetasyon indeksleri (12 indeks)
Temel bileşenler	İlk üç temel bileşen
Ton	(R-G-B; NIR1-R-G; NIR2-R-G)
Doğunluk	(R-G-B; NIR1-R-G; NIR2-R-G)
Yoğunluk	(R-G-B; NIR1-R-G; NIR2-R-G)
Ortalama	WV-2 (0°, 45°, 90°, 135°)
Varyans	WV-2 (0°, 45°, 90°, 135°)
Homojenlik	WV-2 (0°, 45°, 90°, 135°)
Zıtlık	WV-2 (0°, 45°, 90°, 135°)
Farklılık	WV-2 (0°, 45°, 90°, 135°)
Entropi	WV-2 (0°, 45°, 90°, 135°)
İkinci moment	WV-2 (0°, 45°, 90°, 135°)
Korelasyon	WV-2 (0°, 45°, 90°, 135°)

Uygulamada değerlendirmeye alınan yüksek boyutlu veri seti içerisinde sınıflar arasındaki ayrımın en etkili şekilde gerçekleştirilmesine olanak sağlayan bantların tespit edilmesi ve gereksiz/faydasız bilgi içeren bantların elenerek boyut indirilmesi amacıyla özellik seçimi yoluna gidilmiştir. Özellik seçimi işleminin gerçekleştirilmesinde özyinelemeli özellik eleme esasına dayanan destek vektör makineleri (DVM-ÖÖE) algoritması kullanılmıştır. Optimum özellik boyutunun tespit edilmesi amacıyla DVM-ÖÖE tarafından belirlenen sıralamalar dikkate alınarak en iyi 10, 20, 30, 40, 50, 75 ve 90 özelliği içerecek şekilde 7 farklı veri seti oluşturulmuştur. Farklı boyutlardaki veri setleri sınıflandırma işlemine tabi tutularak özellik boyutundaki artışın sınıflandırma doğruluğuna olan etkileri incelenmiştir.

Özellik seçimi sonuçlarına göre farklı boyutlarda oluşturulan eğitim veri setleri için DVM ve EYK algoritmaları tarafından elde edilen sınıflandırma modellerinin test verine uygulanması sonucunda elde edilen genel sınıflandırma doğrulukları ve algoritmalarının sınıflandırma performanslarındaki değişimler Şekil 2’de gösterilmiştir. Şekil incelendiğinde, DVM algoritmasının en yüksek sınıflandırma doğruluğuna (%88,23) özellik seçimi sonucunda belirlenen 40 bantı içeren veri setiyle ulaştığı, EYK algoritmasının ise en yüksek sınıflandırma doğruluğuna (%83,74) seçilen 20 banttan oluşan veri seti ile ulaştığı görülmektedir. Farklı boyuttaki veri setleri kullanımında DVM’nin sınıflandırma performansı analiz edildiğinde, sınıflandırma doğruluğunun 40 bantı içeren veri setine kadar artış gösterdiği bu noktadan sonra sınıflandırma performansının düşüş eğiliminde olduğu görülmektedir. Diğer bir ifadeyle, 40 banttan sonra veri setine sonradan dahil edilecek bantlar DVM’nin performansını olumsuz etkilemektedir. EYK algoritması için elde edilen sonuçlar incelendiğinde, 20 bantlı veri setine dahil edilen 10 yeni bantla sınıflandırma sonuçlarında bir miktar azalma olduğunu, bu noktadan sonra sınıflandırma doğruluğunun yaklaşık %3 düşüş göstererek %79 seviyelerinde değerler aldığı görülmektedir. Bu durum EYK algoritmasının veri seti boyutu artışına karşı duyarlı olduğunu ve yüksek boyutlu veri setleri için düşük sınıflandırma performansı sergilediğini göstermektedir.

DVM ve EYK algoritmaları için elde edilen sınıflandırma doğrulukları karşılaştırıldığında, Şekil 2’den de görüleceği üzere tüm sınıflandırma senaryoları için DVM algoritmasıyla EYK yönteminden daha yüksek sınıflandırma sonuçları elde edildiği görülmektedir. En düşük boyutlu veri seti (10 bant) için DVM ve EYK arasındaki doğruluk farkı yaklaşık %7 iken, özellik seçimi yapılmaksızın tüm veri setinin (96 bant) kullanıldığı durumda ise yaklaşık %4 olduğu görülmektedir. Özellikle 40 bantlı veri seti ve daha yüksek boyutlu veri setleri kullanıldığında DVM algoritmasının yaklaşık %9 daha yüksek sınıflandırma doğruluğu ürettiği görülmektedir.

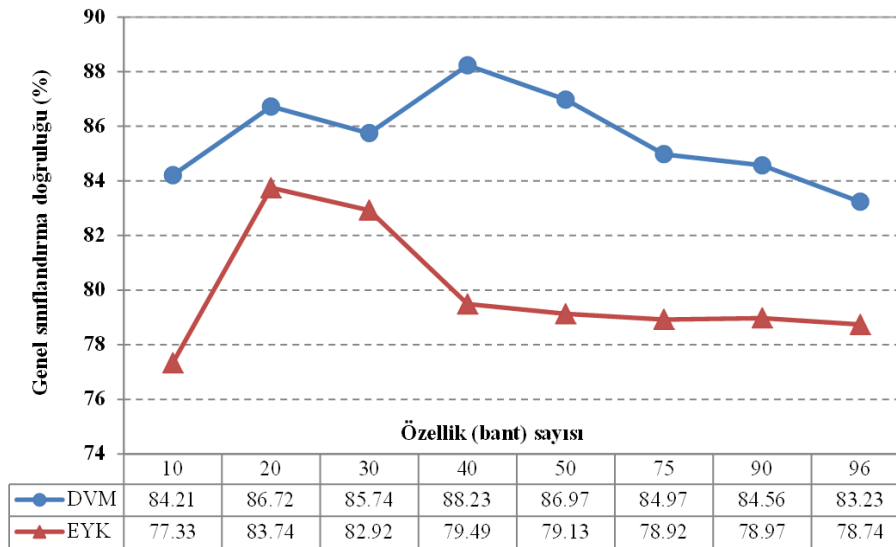
En yüksek sınıflandırma doğruluğunun elde edildiği 40 bantlı veri seti analiz edildiğinde, veri setinin WV-2 görüntüsünün 8 spektral bantını, ilk iki temel bileşeni, 9 vejetasyon indeksini, 14 doku özelliğini ve 7 ton-doygunluk-yoğunluk dönüşümü bileşeninden oluştuğu görülmüştür. Veri seti içerisindeki özellikler (bantlar) açısından ele alındığında, WV-2

görüntüsünün kırmızı, yeşil ve kızılötesi-2 bantlarının ön plana çıktığı görülmektedir. Bunun temel nedeni, söz konusu bantların spektral yansıma değerlerinin yanında ve bu bantlar yardımıyla hesaplanan ton-doygunluk-yoğunluk bileşenlerinin ve farklı doku özelliklerinin seçilen 40 bant içerisinde yer almasıdır. Uygulamada değerlendirmeye alınan vejetasyon indeksleri içerisinde geliştirilmiş bitki örtüsü indeksi, atmosferik koşullara dayanıklı bitki indeksi ve temel bileşen-kızılötesi bant yarımıyla normalleştirilmiş indeks haricindeki tüm vejetasyon indekslerinin 40 bantlı veri seti içerisinde yer almıştır. DVM-ÖÖE ile seçilen 40 bant içerisinde doku özelliklerinden ortalama, varyans, zıtlık ve entropi bileşenlerinin bulunduğu tespit edilmiştir.

DVM ve EYK algoritmaları için hesaplanan en yüksek genel doğruluklar arasındaki % 4,49’luk farkın istatistiksel anlamlılığı McNemar’s testi kullanılarak analiz edilmiştir. Parametrik olmayan ve ki-kare dağılımını esas alan McNemar’s test istatistiği Eşitlik 1 yardımıyla hesaplanmaktadır.

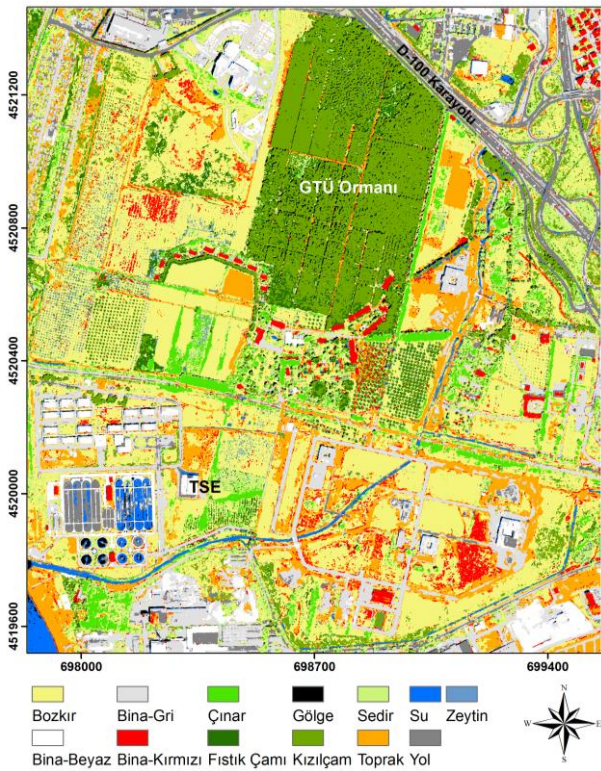
$$McNemar(\chi^2) = \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} \quad (1)$$

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. DVM ve EYK algoritmalarına ilişkin sonuçlara McNemar’s testi uygulandığında hesaplanan istatistik değeri 14,56’dır. Bu durum DVM ve EYK algoritmaları için söz konusu sınıflandırma performansı farklılığının istatistiksel olarak anlamlı bir fark olduğunu, dolayısıyla iki sınıflandırma sonucunun birbirinden farklı olduğunu göstermektedir.



Şekil 2. Farklı sayıda bant kombinasyonları için sınıflandırma performanslarındaki değişimler.

Çalışma alanına ait tematik harita üretiminde en yüksek sınıflandırma doğruluğunun elde edildiği sınıflandırma modeli (DVM-ÖÖE ile seçilen 40 bantlı veri seti ve DVM sınıflandırıcısı) kullanılmıştır. Söz konusu model ile tematik harita üretimi için 96 bantlı tüm veri seti DVM-ÖÖE ile seçilen 40 bantlı içerecek şekilde yeniden düzenlenerek sınıflandırma işlemine tabi tutulmuştur. Sınıflandırma sonucunda üretilen tematik harita Şekil 3'te gösterilmiştir. Harita incelendiğinde bazı arazi örtüsü ve kullanım sınıfları arasında hatalı sınıflandırmalar olduğu görülmektedir. Özellikle toprak sınıfına ait pikseller kiremit malzemeye kaplı bina sınıfını temsil eden bina-kırmızı sınıfına atanarak hatalı sınıflandırılmıştır. Bu durumun temel nedeni olarak iki sınıf arasındaki spektral benzerlik olduğu düşünülmektedir. Diğer taraftan çalışma alanı içerisindeki geniş ve iğne yapraklı türlerin büyük ölçüde doğru sınıflandırıldığı, zeytin ağacını temsil eden piksellerin bir kısmının canlı vejetasyonu temsil eden bozkır sınıfı içerisindeki piksellerle karıştığı görülmektedir. Çalışma alanının belirli kısımlarında benzer spektral özelliklere sahip su geçirimsiz yüzeyler olarak tanımlanan yol ve bina-gri sınıfına ait pikseller arasında da hatalı sınıflandırmalar olduğu tespit edilmiştir.



Şekil 3. DVM-ÖÖE tarafından seçilen 40 bantlı veri setinin DVM ile sınıflandırılması sonucunda üretilen tematik harita.

7. SONUÇLAR

Bu çalışmada yüksek çözünürlüklü WV-2 görüntüsünün yardımıyla birçoğu spektral benzerlik gösteren 13 farklı arazi kullanımı ve arazi örtüsü sınıfının ayırt edilmesi ve tematik harita üretilmesi problemi ele alınmıştır. Sınıfları temsil eden piksellerin kendi içerisindeki ve birbirleri arasındaki ayrımın artırılması amacıyla farklı vejetasyon indeksleri, doku özellikleri, temel bileşenler analizi ve ton-doygunluk-yoğunluk dönüşümü neticesinde hesaplanan bileşenler yardımcı veri seti olarak tanımlanmıştır. WV-2 ve yardımcı verilerin birleşimi ile oluşan 96 bantlı veri seti içerisindeki optimum bantların

belirlenmesinde özyinelemeli özellik eleme esasına dayanan destek vektör makineleri (DVM-ÖÖE) algoritması kullanılmıştır. Optimum veri seti boyutunun tespiti amacıyla özellik seçimi sonuçları göz önüne alınarak farklı boyutlarda veri setleri oluşturulmuş ve söz konusu veri setleri destek vektör makineleri (DVM) ve en yakın komşuluk (EYK) algoritması kullanılarak sınıflandırılmıştır.

Çalışma sonuçları analiz edildiğinde farklı boyutlarda üretilen veri setlerinin tamamı için DVM'nin EYK algoritmasından %8'lere varan düzeylerde daha doğru sonuçlar ürettiği görülmüştür. Bu durum sınırlı sayıda spektral veri içeren veri setleri olması durumunda veya yüksek boyutlu veri seti kullanımında DVM'nin klasik EYK algoritmasına göre daha üstün bir performans sergilediğini destekler niteliktedir. Veri seti boyutuna göre her bir sınıflandırıcının performansı ayrı ayrı analiz edildiğinde, DVM 40 bantlı veri setiyle EYK algoritması ise 20 bantlı veri seti ile en yüksek sınıflandırma doğruluğuna ulaşmıştır. Algoritmalar için tespit edilen bu optimum bant sayısından sonra veri setine dahil edilecek yeni bantların algoritmaların performansına katkısının olmadığı hatta doğruluğun azaldığı görülmüştür. Elde edilen bu bulgu, her iki algoritmanın da Hughes fenomeninden veya boyutsallık probleminde etkilendiğini göstermektedir.

Özellik seçiminin sınıflandırma algoritmalarının performanslarına olan etkileri incelendiğinde, seçilen ve boyutu azaltılan veri seti ile DVM'nin sınıflandırma performansında %5 artış görülürken, EYK algoritmasının performansındaki iyileşmenin %6 olduğu tespit edilmiştir. Elde edilen bu sonuç, DVM-ÖÖE algoritması kullanılarak gerçekleştirilen özellik seçimiyle veri seti içerisindeki gereksiz/faydasız bilgi içeren bantların sayısı %60-80 oranında azaltılarak sınıflandırma performansında %5-6 seviyelerinde artış olduğu şeklinde ifade edilebilir.

Çalışma sonucu elde edilen bulgular, yüksek çözünürlüklü uydu görüntüleri yardımıyla arazi örtüsü veya kullanımının tespiti probleminde yardımcı veri seti kullanımının olumlu etkileri olduğunu, özellikle özellik seçimi yaklaşımıyla sınıflandırma doğruluğunda önemli değişimler olabileceğini göstermektedir. Bununla birlikte çalışma sonuçları, birçok sınıflandırma ve örüntü tanımlama probleminde başarıyla kullanılan DVM algoritmasının, yüksek boyutlu veri seti içerisinde sınıflandırma probleminin çözümüne katkı sağlayacak optimum bantların tespit edilmesinde de başarılı bir algoritma olduğunu desteklemektedir. Diğer taraftan üretilen tematik harita doğruluğunun düşük olması ve haritada üzerinde tespit edilen hatalı sınıflandırmaların piksel tabanlı yaklaşımın bir sonucu olduğu düşünülmekte, sonraki çalışmalar için obje tabanlı yaklaşımın kullanılmasının uygun olacağı değerlendirilmektedir.

KAYNAKLAR

- Atkinson, J. T., Ismail, R., Robertson, M., 2014. Mapping Bugweed (*Solanum mauritianum*) Infestations in *Pinus patula* Plantations Using Hyperspectral Imagery and Support Vector Machines. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7 (1), pp.17-28.
- Baccianella, S., Esuli, A., Sebastiani, F., 2014. Feature Selection for Ordinal Text Classification. *Neural Computation*, 26(3), pp.557-591.

- Bannari, A., Morin, D., Bonn, F., 1995. A review of vegetation indices, *Remote Sensing Reviews*, 13(1-2), 95-120.
- Campbell, J.B and Wynne, R.H., 2011. *Introduction to Remote Sensing* (5th ed.), New York, USA: The Guilford Press.
- Carleer, A. P., Debeir, O., Wolff, E., 2005. Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering and Remote Sensing*, 71 (11), pp.1285-94.
- Çölkesen, İ., Yomralıoğlu, T., 2014. Arazi örtüsü ve kullanımının haritalanmasında WorldView-2 uydu görüntüsü ve yardımcı verilerin kullanımı, *Harita Dergisi*, 152, pp.12-24.
- Fassnacht, F. E., Neumann, C., Forster, M., Buddenbaum, H., Ghosh, A., Clasen, A., Joshi, P. K., Koch, B., 2014. Comparison of Feature Reduction Algorithms for Classifying Tree Species With Hyperspectral Data on Three Central European Test Sites. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6), pp.2547-2561.
- Foody, G. M., 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 70(5), pp.627-633.
- 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, pp.298-311.
- Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1-3), pp.389-422.
- Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6), pp.610-621.
- Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), pp.725-749
- Hughes, G.F., 1968. On the mean accuracy of statistical pattern recognizers, *IEEE Transactions on Information Theory*, 14(1), pp.55-63.
- Kavzoglu, T., Colkesen, I., 2009. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, 11(5), pp.352-359.
- Kavzoglu, T., Mather, P.M., 2000. The Use of Feature Selection Techniques in the Context of Artificial Neural Networks. *Proceedings of the 26th Annual Conference of the Remote Sensing Society* (CD-ROM), Leicester, UK.
- Lillesand, T.M., Kiefer, R.W., Chipman, J.W., 2007. *Remote Sensing and Image Interpretation*. Sixth edition, New York, USA: John Wiley & Sons.
- Low, F., Michel, U., Dech, S., Conrad, C., 2013. Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. *ISPRS Journal of Photogrammetry and Remote Sensing*, 85, pp.102-119.
- 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), pp.1159-1168.
- Mather, P.M., Koch, M. 2011. *Computer Processing of Remotely-Sensed Images: An Introduction*. Fourth Edition. Chichester, UK: Wiley-Blackwell.
- Mianji, F. A., Zhang, Y., 2011. Robust Hyperspectral Classification Using Relevance Vector Machine. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6), pp.2100-2112.
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), pp.247-259.
- Mwangi, B., Tian, T. S., Soares, J. C., 2014. A Review of Feature Reduction Techniques in Neuroimaging. *Neuroinformatics*, 12(2), pp.229-244.
- Pal, M., Foody, G. M., 2010. Feature Selection for Classification of Hyperspectral Data by SVM. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5), pp.2297-2307.
- Payero, J. O., Neale, C. M. U., Wright, J. L., 2004. Comparison of eleven vegetation indices for estimating plant height of alfalfa and grass. *Applied Engineering in Agriculture*, 20(3), pp.385-393.
- Rencz, A.N., Ryerson, R.A., 1999). *Manual of Remote Sensing*, Remote Sensing for the Earth Sciences. New York, USA: John Wiley & Sons.
- Richards, J.A., 2013. *Remote Sensing Digital Image Analysis: An Introduction*. Fifth Edition. New York, USA: Springer-Verlag.
- Shamsoddini, A., Trinder, J. C., Turner, R., 2013. Pine plantation structure mapping using WorldView-2 multispectral image. *International Journal of Remote Sensing*, 34(11), pp.3986-4007.
- Vapnik, V. N., 1995. *The nature of statistical learning theory*. New York, USA: Springer-Verlag.
- 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), pp.4515-4545.
- Zhang, R., Ma, J. W., 2009. Feature selection for hyperspectral data based on recursive support vector machines. *International Journal of Remote Sensing*, 30 (14), pp.3669-77.