

[970]

NESNE-TABANLI SINIFLANDIRMADA FİLTRELEME TABANLI ÖZELLİK SEÇİMİ ALGORİTMALARININ KULLANIMI VE SINIFLANDIRMA DOĞRULUĞUNA ETKİLERİNİN İNCELENMESİ

İsmail ÇÖLKESEN¹, Taşkın KAVZOĞLU²

¹Yrd.Doç.Dr., Gebze Teknik Üniversitesi, Jeodezi ve Fotogrametri Mühendisliği Bölümü, 41400, Gebze, Kocaeli, icolkesen@gtu.edu.tr

²Prof. Dr., Gebze Teknik Üniversitesi, Jeodezi ve Fotogrametri Mühendisliği Bölümü, 41400, Gebze, Kocaeli, kavzoglu@gtu.edu.tr

ÖZET

Yüksek çözünürlüklü görüntülerin miktarındaki artışla birlikte nesne-tabanlı görüntü analizi yaklaşımı, detaylı arazi kullanımı ve arazi örtüsü haritalarının üretiminde büyük önem kazanmıştır. Segmente edilen görüntü objeleri için birçok özelliğin tanımlanması nesne-tabanlı görüntü analizinin en önemli avantajlarından biridir. Buna karşın fazla sayıda özelliğin kullanımı ihtiyaç duyulan hesaplama zamanını artırmakta ve sınıflandırma doğruluğunda azalmaya neden olabilmektedir. Bu nedenlerden dolayı özellik seçimi nesne-tabanlı görüntü analizinde önemli bir çalışma konusu haline gelmiştir. Bu çalışmada, WorldView-2 görüntüsünün kullanarak çoklu çözünürlük segmentasyonu ile üretilen görüntü nesnelere için tanımlanan 110 özellik içerisinde en etkili olan özelliklerin tespit edilmesi için filtreleme tabanlı özellik seçimi algoritmalarından ki-kare, bilgi kazancı ve ReliefF algoritmaları kullanılmıştır. Bu amaç doğrultusunda, öncelikle filtreleme algoritmaları kullanılarak 110 nesne özelliği için önem dereceleri belirlenmiştir. Daha sonra, korelasyon tabanlı fayda fonksiyonu kullanılarak optimum özellik sayısı tespit edilmiştir. Seçilen özelliklere sahip görüntü nesnelere sınıflandırılmasında en yakın komşuluk, destek vektör makineleri ve rastgele orman sınıflandırıcılarından yararlanılmıştır. Çalışma sonuçları değerlendirmeye alınan tüm özellik seçimi algoritmalarının yüksek boyutlu veri seti boyutunun azaltılmasında etkili olduğunu göstermiştir. Tüm sınıflandırma algoritmaları en yüksek sınıflandırma doğruluklarına bilgi kazancı tarafından 110 özellik içerisinde seçilen 24 özellik kullanılarak ulaşmıştır. Bilgi kazancı algoritması kullanımıyla sınıflandırma doğruluğundaki artış %4 seviyelerine ulaşmış ve Wilcoxon işaretli sıralar testi sonuçları göre performans farklılıklarının istatistiksel olarak anlamlı olduğunu göstermiştir.

Anahtar Sözcükler: Nesne-tabanlı sınıflandırma, Destek vektör makineleri, Rastgele orman, Ki-kare, Bilgi kazancı, ReliefF

ABSTRACT

THE USE OF FILTER BASED FEATURE SELECTION ALGORITHMS IN OBJECT-BASED CLASSIFICATION AND INVESTIGATION OF THEIR EFFECTS ON CLASSIFICATION ACCURACY

With the availability of high resolution satellite image, object-based image analysis has gained considerable importance for producing detailed land use and land cover maps. One of the main advantages of the object-based image analysis is that it allows variety of features to be identified for the segmented image objects. However, the use of a large number of features increases the required computational time and results in decrease in classification accuracy. For these reasons, feature selection has become an important research topic in object-based image analysis. In this study, three filter-based feature selection algorithm namely, chi-square, information gain and ReliefF algorithms were evaluated to determine effective features within the available 110 features identified for the image objects created by multi-resolution segmentation algorithm using the high resolution WorldView-2 image. For this purpose, importance degree (ranks) of 110 input object features were firstly determined by filtering algorithms. Then, correlation-based merit function was utilized to determine optimum feature subset size according to the ranked order of the features. Nearest neighbour, support vector machines and random forest classifiers were utilized to classify segmented image objects having selected object features. Results of this study showed that all considered feature selection algorithms were found to be effective for reducing high dimensional data sets. In particular, the algorithms reached to their best performances with 24 selected features within the available 110 input features by information gain algorithm. The improvement in classification accuracy reached to 4% and the performance differences were observed to be statistically significant considering the Wilcoxon signed ranks test results.

Keywords: Object-based classification, Support vector machines, Random Forest, Chi-square, Information gain, ReliefF.

1.GİRİŞ

Yüksek çözünürlüklü uydu görüntülerinin kullanımının artmasıyla birlikte tematik harita üretiminde klasik piksel-tabanlı sınıflandırma yaklaşımı yerine nesne-tabanlı yaklaşımın kullanımı ön plana çıkmış ve yöntemin sınıflandırma doğruluğu üzerindeki olumlu etkileri ortaya koyulmuştur (Myint et al., 2011; Duro et al., 2012; Demers et al., 2015). Nesne-tabanlı yaklaşım ile sınıflandırma işleminin gerçekleştirilmesinde görüntü pikselleri yerine segmentasyon sonucu üretilen görüntü nesnelere kullanılmaktadır. Nesne-tabanlı sınıflandırma işleminde, oluşturulan görüntü nesnelere ilişkin spektral, dokusal ve içeriksel birçok özellik tanımlanabilmekte ve

sınıflandırmada bu özelliklerden yararlanabilmektedir. Nesnelere ilişkin söz konusu özelliklerin tanımlanması neticesinde elde edilen veri seti bir biri ile korelasyonlu, tekrarlı veriyi içerir duruma gelmektedir. Bu nedenle, nesne-tabanlı sınıflandırma işleminde özellik seçimi ile veri seti boyutunun azaltılması ve sınıflandırma işlemine katkı sağlayacak nesne özelliklerinin tespit edilmesi güncel bir araştırma konusudur.

Bu çalışmada temel olarak WorldView-2 görüntüsünün nesne-tabanlı yaklaşımla sınıflandırılması ve arazi örtüsü/kullanımını gösteren tematik harita üretimi problemi ele alınmış ve özellik seçiminin sınıflandırma doğruluğu üzerindeki etkisi araştırılmıştır. Segmentasyon sonucu elde edilen görüntü nesnelere ilişkin tanımlanan özellikler içerisinde sınıflandırma probleminin çözümüne katkı sağlayacak en etkili özelliklerin tespit edilmesinde filtreleme tabanlı özellik seçimi algoritmalarından ki-kare, kazanç oranı ve Relief algoritmalarından yararlanılmıştır. Seçilen özellikleri içeren veri setlerinin sınıflandırılması ve tematik harita üretiminde kernel tabanlı destek vektör makineleri (DVM), karar ağaçlarını esas alan rastgele orman (RO) algoritması ve klasik en yakın komşuluk (EYK) algoritması kullanılmış ve özellik seçimi işleminin sınıflandırma doğruluğuna etkileri karşılaştırılmalı olarak analiz edilmiştir. Elde edilen sınıflandırma doğruluklarının karşılaştırılmasının yanında veri seti boyutunun azalması ile elde edilen doğruluklar arasındaki farklılıkların istatistiksel anlamlılığı Wilcoxon işaretli sıralar testi kullanılarak analiz edilmiştir.

2.ÇALIŞMA ALANI VE KULLANILAN VERİ

Yüksek çözünürlüklü uydu görüntüsü üzerinden obje-tabanlı görüntü analizi yardımıyla arazi örtüsü/kullanım türlerini temsil eden tematik harita üretimi amacıyla Kocaeli iline bağlı Darıca ilçesini içerisine alan yaklaşık 280 hektarlık bölge çalışma alanı olarak belirlenmiştir (Şekil 1). İlçe mavi bayraklı plajları, kumsalları, çeşitli rekreasyon alanları ve birçoğu yazlık evlerden oluşan yapısıyla Marmara bölgesinin en önemli turizm ve cazibe merkezlerindedir. Çalışma alanı içerisinde 10 farklı arazi örtüsü ve arazi kullanım sınıfı olduğu belirlenmiştir. Bunlar içerisinde su, toprak, bozkır, asfalt zeminli yollar, geniş ve iğne yapraklı ağaçlardan oluşan temel arazi örtüsü sınıflarına ilave olarak, bölgedeki binalar, çatı yapımında kullanılan materyallere göre kırmızı, gri ve beton olarak üç farklı sınıf altında ele alınmıştır. Ayrıca, gölge probleminden kaynaklanabilecek sınıflandırma hatalarının azaltılması amacıyla gölge sınıfı da tanımlanarak uygulamada dikkate alınmıştır. Sınıflandırma ve tematik harita üretiminde çalışma alanını içerisine alan 2013 tarihinde kaydedilmiş WorldView-2 (WV-2) uydu görüntüsünden yararlanılmıştır. Uygulama kapsamında, nesne-tabanlı görüntü analizinin gerçekleştirilmesinde, WV-2 uydu görüntüsünün 2m yersel çözünürlüğe sahip 8 multispektral bandı ile 0,5m çözünürlükteki pankromatik bandının kaynaştırılması neticesinde elde edilen kaynaştırılmış görüntü kullanılmıştır.



Şekil 1. Çalışma alanı ve konumu

3.NESNE TABANLI GÖRÜNTÜ ANALİZİ

3.1.Görüntü segmentasyonu

Nesne tabanlı görüntü analizinin esasları benzer spektral özelliklere sahip piksellerin belirli kurallara çerçevesinde bir araya getirilerek homojen yapıları görüntü nesnelere elde edilmesi (görüntü segmentasyonu) ve söz konusu görüntü nesnelere ilişkin spektral, geometrik ve içeriksel özellikler yardımıyla söz konusu nesnelere belirli bir arazi örtüsü veya kullanım sınıfına atanması esastır. Bu açıdan ele alındığında nesne tabanlı görüntü analizinde, görüntü segmentasyonu, görüntü nesnelere için özellik tanımlama (özellik seçimi) ve tanımlanan özellikler

yardımla görüntü nesnelere belirli bir arazi örtüsü veya arazi kullanım sınıflarına atanması olarak bilinen üç temel işlem adımı mevcuttur (Belgiu et al., 2014). Nesne tabanlı sınıflandırma yaklaşımında sınıflandırmaya esas olan birim tek bir piksel değil birden çok pikselin bir araya gelmesiyle elde edilen görüntü nesnelere aittir. Diğer bir ifadeyle nesne tabanlı yaklaşımda milyonlarca piksellerden oluşan görüntünün değerlendirilmesi yerine bu pikselleri temsil eden nesnelere değerlendirilmesi söz konusudur. Bu nedenle nesne tabanlı görüntü analizinde, segmentasyon işlemi en temel işlem adımı olup, segmentasyon neticesinde üretilen görüntü nesnelere kalitesi tematik harita doğruluğuna etki eden önemli bir bileşendir. Günümüze kadar segmentasyon işleminin gerçekleştirilmesi amacıyla satranç tahtası, dörtlü ağaç, spektral farklılık, kontrast ayırma, çok eşik segmentasyonu ve çoklu çözünürlüklü segmentasyon gibi çok çeşitli yaklaşımlar geliştirilmiş ve uygulamalarda kullanılmıştır (eCognition Developer 9.0). Bunlar arasında çoklu çözünürlüklü segmentasyonu literatürde en çok kullanılan ve popüler segmentasyon algoritmasıdır. Algoritma iteratif bir yaklaşımla görüntü nesnelere elde edilmesini esas alır ve işlem başlangıcında her bir görüntü pikseli bir görüntü nesnesi olarak değerlendirilir (Baatz and Schape, 2000). Çoklu çözünürlük segmentasyonu ile görüntü nesnelere elde edilmesinde ölçek, şekil ve yoğunluk olarak adlandırılan üç temel parametrenin kullanıcı tarafından belirlenmesi gerekmektedir. Bunlardan en önemlisi segmentasyon ile oluşturulacak görüntü nesnelere boyutunu belirleyen ölçek parametresi olarak ifade edilmektedir (Blaschke 2010; Myint et al, 2011; Belgiu et al., 2014).

3.2.Nesne özelliklerinin tanımlanması ve özellik seçimi

Nesne tabanlı görüntü analizinde segmentasyon sonucu elde edilen nesnelere ilişkin özelliklerin tanımlanması ikinci işlem adımıdır. Sınıflandırma işleminde sadece parlaklık değerlerinin değil nesnelere ilişkin tanımlanabilen birçok özelliğin kullanılabilir olması nesne tabanlı yaklaşımın en önemli avantajlarından biridir. Kullanılan nesne özellikleri arasında nesnelere ilişkin piksellerin temel spektral özellikleri (ortalama değer, standart sapma vb.), nesnelere ilişkin doku özellikleri ve nesnelere ilişkin geometrik/şekilsel özellikleri gibi birçok özellik yer almaktadır. Tanımlanan özellikler yardımla ilgili duyulan arazi örtüsü ve arazi kullanım sınıflarının birbirinden ayrt edilmesi noktasında önemli bilgiler elde edilebilmektedir. Ancak yüzlerce özelliğin bir arada değerlendirilmesi sadece piksellere ilişkin yansıma değerlerinin kullanıldığı piksel tabanlı yaklaşıma göre hesaplama için ihtiyaç duyulan zamanı arttırmaktadır. Bunun yanında yüksek sayıda özellik kullanımı ile birlikte oluşturulacak sınıflandırma modeli daha karmaşık bir hale gelerek, literatürde Hughes fenomeni yada boyutsallık problemi olarak bilinen ve sınıflandırma doğruluğunu olumsuz etkileyen sonuçların ortaya çıkmasına neden olabilmektedir (Kavzoglu and Mather, 2002; Pal and Foody, 2010, Ma et al., 2015). Dolayısıyla gereksiz veya tekrarlı bilgiler içeren özelliklerin elenerek faydalı bilgiler taşıyan özelliklerin seçilmesi yoluyla gereksiz bellek ihtiyacının önüne geçilerek sınıflandırma işleminde kullanılan algoritmaların karmaşık yapısı sadeleştirilebilmektedir (Camps-Valls et al., 2011). Bu nedenle özellik seçimi nesne tabanlı görüntü analizi yardımla gerçekleştirilecek sınıflandırmanın doğruluğunu ve etkinliğini arttıracak önemli bir işlem adımıdır. Özellik seçiminin temel amacı sınıflar arasındaki ayrımı kolaylaştırarak sınıflandırma algoritmasının tahmin kabiliyetini güçlendirebilecek özelliklerin belirlenmesi ve fayda sağlamayan özelliklerin elenerek veri seti boyutunun azaltılmasıdır (Guyon and Elisseeff, 2003).

Özellik seçimi başta uzaktan algılama olmak üzere, veri madenciliği, metin ve görüntü işleme uygulamalarında yaygın kullanıma sahip bir tekniktir (Baccianella et al., 2014; Fassnacht et al., 2014; Mwangi et al., 2014). Bu amaca yönelik olarak literatürde birçok metod önerilmiştir (Saeys ve diğ, 2007; Kumar ve Minz, 2014). Özellik seçiminde filtreleme (Blum ve Langey, 1997) ve sarmalama (wrappers) (Kohavi ve John, 1997) olarak bilinen iki temel yaklaşım kullanılmaktadır. Filtreleme metodları özellik seçiminde dolaylı ölçütler kullanılmaktadır. Bu ölçütlere örnek olarak özellikler arasındaki korelasyonun değerlendirilmesi veya veri setindeki her bir özelliğin sınıflandırma doğruluğuna etkisinin değerlendirilmesi gösterilebilir. Filtre yönteminde özellik seçiminde kullanılan algoritmanın performansının optimize edilmesinden ziyade bir ölçüt veya değerlendirme fonksiyonunu maksimum yapan alt küme belirlenir. Ayrıca, sınıflandırma sonucunda ulaşılmak istenen hedef sınıflar ve seçilecek bantlar arasında güçlü bir ilişki olduğu göz önüne alınarak sınıfları birbirinden en iyi şekilde ayrt edebilecek özelliklerin seçilmesi söz konusudur. Bu nedenle filtre yönteminde belirli bir sınıflandırma probleminin çözümünde değerlendirmeye alınan her bir özelliğin problem çözümündeki katkısı veya önem derecesi belirlenir. Literatürde filtreleme yöntemiyle özellik seçiminde değerlendirme ölçütü olarak birçok algoritma geliştirilmiştir. Bunlar arasında en popüler metodlar olarak ki-kare istatistiğine dayanan ki-kare özellik seçimi algoritması, karar ağaçları sınıflandırıcısında entropiye dayalı olarak ağaç yapısındaki dallanmayı belirleyen bilgi kazancı ve en yakın komşuluk algoritmasını esas alan Relief algoritması verilebilir.

Ki-kare testi (χ^2) iki özellik arasındaki ilişkinin bağımlı veya bağımsız olduğunu belirlemeye yarayan ve her bir özelliğin sınıf etiketlerine göre birbirinden bağımsız olarak değerlendirilen bir algoritmadır (Plackett, 1983; Mingers, 1989). Ki-kare testi her bir özellikteki sınıflara ait değerlerin dağılımını inceler. Hesaplanan istatistik değeri arttıkça değerlendirmeye alınan özellikler sınıflar arasında yüksek derecede bağımlılık olduğu anlamına gelmektedir. Diğer bir ifadeyle değerlendirmeye alınan özelliğin sınıflar ile ilişkili olduğu ifade edilir (Çölkesen, 2015). Bilgi kazancı makine öğrenmesi ve bilgi teorisi alanlarında yaygın olarak kullanılan bir ölçüttür (Quinlan, 1993). Entropi

kavramını esas alan algoritma temel olarak veri setini oluşturan her bir özelliğin sınıfları ayırt etme potansiyelini ölçmektedir. Örneğin nesnelere ilişkin tanımlanan standart sapma özelliği (*Std*) ve toprak sınıfı (*T*) için bilgi kazancı hesaplandığı göz önüne alınırsa, bilgi kazancı *Std* bilindiğinde *T* değerindeki belirsizliğin azaltılmasıdır. Diğer bir ifadeyle hesaplanan bilgi kazancı değeri yüksek ise söz konusu özellik yardımıyla ilgili sınıfa ilişkin daha çok bilgi edinilebileceği sonucuna varılır. Her bir özellik için bu işlem iteratif olarak tekrar edilerek tüm özelliklere ait bilgi kazancı değerleri veya puanlamalar hesaplanır. Diğer bir filtreleme algoritması olan Relief algoritması özellik seçiminde en yakın komşuluk algoritmasına benzer bir değerlendirme yapmaktadır (Kira and Rendell, 1992). Algoritmanın temel çalışma prensibi, veri seti içerisindeki her bir özelliğin benzer yapıdaki görüntü nesnelere ayırt edebilme kabiliyetinin ölçülmesi esasına dayanmaktadır. Bu noktada belirli bir mesafeye yakın uzaklıkların aynı sınıfa ait olma olasılığı yüksektir prensibi göz önüne alınmaktadır (Kononenko, 1994). Söz konusu algoritmalar kullandıkları farklı ölçütlere göre özellikleri önem derecelerine göre sıralama yaparken, en yüksek dereceye sahip kaç tane özellik ile sonuca gidilebileceği bilgisini vermemektedir. Diğer bir ifadeyle özellik sayısını azaltıp, sınıflar arasında en iyi ayrımı yapan özellikleri içeren alt küme oluşturamamaktadır. Bu problemin çözümü için uygulama kapsamında korelasyon tabanlı fayda fonksiyon kullanımı önerilmiştir. Söz konusu fonksiyon en iyi alt kümeyle dahil edilecek özelliklerin birbirleriyle korelasyonuz fakat sınıflar ile yüksek korelasyonlu olması prensibine dayanmaktadır (Hall, 1998). Söz konusu fayda fonksiyonu,

$$r_{zc} = \frac{\overline{kr_{zi}}}{\sqrt{k + k(k-1)r_{ii}}} \quad (1)$$

eşitliği ile hesaplanmaktadır. Bu eşitlikte r_{zc} özellikler ve sınıflar arasındaki toplam korelasyonu, k özellik sayısını, $\overline{r_{zi}}$ özellikler ile sınıflar arasındaki ortalama korelasyonu ve $\overline{r_{ii}}$ ise özellikler arasındaki korelasyonun ortalamasını göstermektedir.

Sınıflandırma Metotları

Bu çalışmada filtreleme algoritmaları ile seçilen özelliklerden oluşan farklı boyutlardaki veri setlerinin sınıflandırılması ve tematik harita üretiminde en yakın komşuluk, rastgele orman ve destek vektör makineleri sınıflandırma algoritmaları kullanılmıştır. 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.

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. Bu çalışmada kernel fonksiyonu olarak sınıflandırma problemindeki etkinliği nedeniyle radyal tabanlı fonksiyon tercih edilmiştir (Kavzoglu and Colkesen, 2009). 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 yapıyla 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 veya görüntü nesnesi), tüm ağaç tahminlerinde en fazla oy alan sınıfa atanması suretiyle sınıflandırılır.

4.UYGULAMA

Bu çalışmada yüksek çözünürlüklü WV-2 uydu görüntüsünün nesne tabanlı görüntü analizi yardımıyla sınıflandırılması ve çalışma alanı içerisinde tespit edilen arazi örtüsü ve arazi kullanım sınıflarını temsil eden tematik harita üretilmesi problemi ele alınmıştır. Bu problemin çözümünde, görüntü nesnelere ilişkin tanımlanan özelliklerin tamamının kullanımı yerine filtreleme tabanlı özellik seçimi algoritmaları tarafından belirlenen

özelliklerin kullanımı ve özellik seçiminin sınıflandırma doğruluğuna etkileri araştırılmıştır. Bu amaca yönelik olarak nesne tabanlı görüntü analizindeki segmentasyon, nesnelere ilişkin özelliklerin belirlenmesi, filtreleme tabanlı özellik seçimi algoritmaları yardımıyla en uygun nesne özelliklerinin tespiti ve sınıflandırma olarak bilinen işlem adımları ayrı ayrı gerçekleştirilmiştir.

Çalışma kapsamında WV-2 uydu görüntüsün segmente edilmesi ve görüntü nesnelere ilişkin özelliklerin tanımlanmasında eCognition Developer 9.10 yazılımından yararlanılmıştır. Segmentasyon işlemi literatürde yaygın olarak kullanılan çoklu çözünürlük segmentasyonu algoritmasından yararlanılmıştır. Algoritmaya ilişkin parametrelerin tespitinde literatürde yaygın bir kullanıma sahip olan ve kullanıcı tarafından farklı ölçek parametreleri ile üretilen segmentasyonların arazi örtüsü ve kullanım sınıflarının temsili noktasındaki yeterliliğine göre karar verildiği deneme-yanılma yaklaşımı kullanılmıştır (Myint et al. 2011; Duro et al. 2012). Segmentasyon işleminin gerçekleştirilmesinde WV-2 görüntüsünün tüm bantları kullanılmış ve her bantın ağırlığı eşit alınmıştır. Farklı ölçek parametreleri ile gerçekleştirilen segmentasyonlar neticesinde, uygulamada değerlendirmeye alınan veri seti için optimum ölçek parametresinin 20 olduğu tespit edilmiştir. Şekil ve yoğunluk parametreleri için parametre değerleri 0,3 ve 0,5 olarak seçilmiştir. Söz konusu parametre değerleri ile gerçekleştirilen segmentasyon işlemi neticesinde 114.538 görüntü nesnesi elde edilmiştir. Söz konusu görüntü nesnelere, arazi çalışmaları neticesinde belirlenen ve çalışma alanı içerisindeki arazi örtüsü ve arazi kullanım sınıflarını temsil eden örnekleme alanları ile ilişkilendirilerek toplam 1.438 referans görüntü nesnesi belirlenmiştir. Özellik seçimi ve sınıflandırma modellerinin oluşturulması amacıyla elde edilen referans nesnelere içerisinden rastgele örnekleme prensibinden hareketle toplam 989 nesne eğitim ve 449 nesne test veri seti olarak belirlenmiştir.

Nesne tabanlı görüntü analizinde ikinci temel işlem adımı görüntü segmentasyonu sonucunda elde edilen nesnelere ilişkin sınıflandırmaya esas olacak özelliklerin belirlenmesidir. Bu amaca yönelik olarak literatürde nesne tabanlı görüntü analizi yardımıyla tematik harita üretiminde en sık kullanılan spektral ve doku özellikleri değerlendirmeye alınmış ve söz konusu özelliklere ilişkin detaylar Tablo 1’de verilmiştir. Tablodan da görüleceği üzere nesnelere ilişkin piksellerin ortalama, standart sapma, bant oranları, minimum ve maksimum parlaklık değerleri temel spektral özellikler olarak tanımlanmıştır. Bu özelliklere ilave olarak, ton-doygunluk-yoğunluk dönüşümü neticesinde elde edilen bileşenler ve WV-2 görüntüsünün kırmızı, kırmızı kenar, kızılötesi-1 ve kızılötesi-2 bantları yardımıyla hesaplanan 3 farklı normalleştirilmiş vejetasyon indeksi (NDVI) spektral özellik olarak tanımlanmıştır. Spektral özelliklerin yanında, gri düzey eş oluşum matrisi yardımıyla hesaplanan 8 farklı doku özelliği (ortalama, varyans, homojenlik, zıtlık, farklılık, entropi, ikinci moment ve korelasyon) kullanılmıştır. Sonuç olarak segmente edilen görüntü nesnelere için 110 özellik tanımlanmıştır.

Tablo 1. Segmentasyon sonucu elde edilen görüntü nesnelere ilişkin tanımlanan özellikler.

	Nesne özellikleri	Açıklama	Özellik sayısı
Spektral özellikler	Ortalama	WV-2 bantları	8
	Maksimum piksel	WV-2 bantları	8
	Minimum piksel	WV-2 bantları	8
	Standart sapma	WV-2 bantları	8
	Oran	WV-2 bantları	8
	NDVI1	(NIR1-Kırmızı/NIR1+Kırmızı)	1
	NDVI2	(NIR2-Kırmızı/NIR2+Kırmızı)	1
	NDVI3	(NIR2-Kırmızı Kenar/NIR2+Kırmızı Kenar)	1
	Ton	(Kırmızı, Yeşil, Mavi)	1
	Doygunluk	(Kırmızı, Yeşil, Mavi)	1
Yoğunluk	(Kırmızı, Yeşil, Mavi)	1	
Doku özellikleri	Homojenlik	WV-2 bantları (Tüm yönler)	8
	Zıtlık	WV-2 bantları (Tüm yönler)	8
	Farklılık	WV-2 bantları (Tüm yönler)	8
	Entropi	WV-2 bantları (Tüm yönler)	8
	İkinci Moment	WV-2 bantları (Tüm yönler)	8
	Ortalama	WV-2 bantları (Tüm yönler)	8
	Standart sapma	WV-2 bantları (Tüm yönler)	8
	Korelasyon	WV-2 bantları (Tüm yönler)	8

Tanımlanan özellikler içerisinde sınıflandırma probleminin çözümüne katkı sağlayacak özelliklerin tespit edilmesi ve veri seti boyutunun azaltılması amacıyla özellik seçimi yoluna gidilmiştir. Özellik seçimi işleminin gerçekleştirilmesinde filtreleme tabanlı ki-kare, bilgi kazancı ve ReliefF algoritmaları kullanılmış ve oluşturulan eğitim veri setine ayrı ayrı uygulanarak değerlendirmeye alınan özelliklere ait önem dereceleri (puanlamalar) belirlenmiştir. Algoritmalar tarafından belirlenen önem dereceleri dikkate alınarak korelasyon tabanlı fayda fonksiyonu yardımıyla optimum sayıda özelliği içeren alt kümeler ayrı ayrı tespit edilmiştir. Tablo 2’den de görüleceği üzere fayda fonksiyonu sonuçlarına göre ki-kare testinin belirlediği özelliklerden en yüksek dereceye sahip ilk 33 özelliği içeren alt küme ile en yüksek fayda değeri hesaplanmıştır. Aynı şekilde bilgi kazancı algoritmasının seçtiği en yüksek puanlamaya sahip ilk 24 nesne özelliğini içeren alt kümeyle en yüksek fayda değerine ulaşıırken, ReliefF algoritmasının seçtiği 39 nesneyi içeren alt kümeyle optimum fayda hesaplanmıştır. Ki-kare, bilgi kazancı ve ReliefF algoritmalarının seçimleri sonucunda elde edilen alt kümeler içerdiği özellikler açısından analiz edildiğinde, seçilen özelliklerin neredeyse tamamının görüntü nesnelere için tanımlanan spektral özelliklerden oluştuğu görülmektedir (Tablo 2). Spektral özellikler içerisinde WV2 görüntüsünün 8 bandı için hesaplanan ortalama parlaklık değerleri, bant oranlamaları ve NDVI değerlerinin sınıflandırma probleminin çözümü noktasında faydalı bilgiler içerdiği tespit edilmiştir. Diğer taraftan, her üç algoritmada nesnelere için tanımlanan toplam 64 doku özelliği içerisinde WV2’nin kızılötesi-2 bandı kullanılarak hesaplanan standart sapma, kızılötesi-1 bandı için hesaplanan korelasyon ve ortalama doku özelliklerini içerdiği görülmüştür. Diğer bir ifadeyle toplam 64 doku özelliği içerisinde söz konusu üç özellik tüm özellik seçimi algoritmaları tarafından seçilmiştir. Elde edilen sonuçlar uygulamada ele alınan nesne tabanlı sınıflandırma probleminin çözümü noktasında spektral özelliklerin doku özelliklerinden daha faydalı bilgiler içerdiğini göstermektedir.

Tablo 2. Ki-kare, bilgi kazancı ve ReliefF algoritmaları tarafından seçilen özellikler.

		Özellik seçimi algoritmaları				
		Nesne özellikleri	Ki-kare	Bilgi kazancı	ReliefF	Tüm veri seti
Spektral özellikler	Ortalama		8	7	8	8
	Maksimum piksel		3	1	8	8
	Minimum piksel		6	2	8	8
	Standart sapma		2	2	2	8
	Oran		8	7	8	8
	NDVI1		1	1	-	1
	NDVI2		1	-	1	1
	NDVI3		1	1	1	1
	Ton		-	-	-	1
	Doygunluk		-	-	-	1
	Yoğunluk		-	-	-	1
	Doku özellikleri	Homojenlik		-	-	-
Zıtlık			-	-	-	8
Farklılık			-	-	-	8
Entropi			-	-	-	8
İkinci Moment			-	-	-	8
Ortalama			1	1	1	8
Standart sapma			1	1	1	8
Korelasyon			1	1	1	8
Toplam			33	24	39	110

Sınıflandırma ve tematik harita üretimi amacıyla ki-kare, bilgi kazancı ve ReliefF algoritmaları tarafından seçilen özellikler dikkate alınarak eğitim, test ve tüm veri seti yeniden düzenlenmiştir. Oluşturulan eğitim veri setleri kullanılarak EYK, DVM ve RO algoritmaları için ayrı ayrı sınıflandırma modelleri üretilmiş ve oluşturulan modellerin performansı test veri seti için analiz edilmiştir. Nesne-tabanlı sınıflandırma sonuçlarının objektif olarak değerlendirilmesi amacıyla Radoux et al. (2011) tarafından önerilen alan-ağırlıklı yöntem uygulanmıştır. Test veri

setindeki nesnelere alanlarının da dikkate alındığı alan-ağırlıklı hata matrisi yardımıyla hesaplanan genel doğruluklar ve Kappa değerleri Tablo 3’de verilmiştir. Genel doğruluklar incelendiğinde ki-kare, bilgi kazancı ve Relief algoritmaları ile seçilen özellikler ile oluşturulan veri setleri içerisinde üç sınıflandırma algoritması da en yüksek doğruluklarına bilgi kazancı algoritması tarafından seçilen 24 özellikle ulaştığı görülmektedir (Tablo 3). Diğer taraftan tüm nesne özelliklerinin herhangi bir özellik seçimine gidilmeksizin sınıflandırma işlemine tabi tutulduğu durumda üç sınıflandırma algoritması da en düşük sınıflandırma performansını göstermiştir. Diğer taraftan ki-kare ve Relief algoritmaları ile seçilen özellikler için sınıflandırma algoritmalarının kendi performanslarının benzer olduğu ve tüm veri seti kullanımına göre sınıflandırma performansında %2 seviyelerinde artış olduğu görülmüştür. Örneğin, EYK algoritması ki-kare ile seçilen 33 özellikle %87,18, Relief ile seçilen 39 özellikle %87,96 doğruluğa ulaşırken tüm özelliklerin kullanıldığı veri setiyle %85,89 genel doğruluğa ulaşılmıştır. Tüm sonuçlar bir arada değerlendirildiğinde bilgi kazancı algoritması yardımıyla veri seti boyutunun yaklaşık olarak %78 oranında azaldığı ve sınıflandırma doğruluğunda EYK ve RO algoritması için %3, DVM algoritması için %4’e varan seviyelerde artış olduğu tespit edilmiştir.

Tablo 3. Özellik seçimi ile üretilen veri setleri kullanımı sonucunda hesaplanan sınıflandırma doğrulukları.

Algoritma	EYK			DVM		RO	
	Özellik sayısı	Genel Doğruluk (%)	Kappa	Genel Doğruluk (%)	Kappa	Genel Doğruluk (%)	Kappa
Ki-kare	33	87,18	0,853	91,55	0,899	90,48	0,891
Bilgi kazancı	24	89,47	0,879	92,00	0,908	91,07	0,898
Relief	39	87,96	0,862	91,16	0,899	89,62	0,881
Tüm özellikler	110	85,89	0,838	87,71	0,859	87,52	0,857

Bilgi kazancı algoritması kullanımı ile ulaşılan sınıflandırma doğrulukları arasındaki farklılıklar Wilcoxon işaretli sıralar testi kullanılarak analiz edilmiş ve sonuçlar Tablo 4’de verilmiştir. Tablodan da görüleceği üzere hesaplanan istatistiksel değerlerin %95 güven aralığında normal dağılım tablo değerinden (1,96) büyük olduğu tespit edilmiş, dolayısıyla söz konusu farklılıkların istatistiksel olarak anlamlı olduğu sonucuna ulaşılmıştır. Diğer bir ifadeyle bilgi kazancı algoritması ile gerçekleştirilen özellik seçimi tüm sınıflandırma algoritmalarının performansında istatistiksel olarak anlamlı artışlar ortaya çıkarmıştır.

Tablo 4. Kullanılan veri setine göre sınıflandırma performanslarının analizi için hesaplan istatistik değerler.

Açıklama	İstatistiksel Karşılaştırma	İstatistik değeri
Tüm veri seti ile karşılaştırma	EYK (24 özellik) – EYK (110 özellik)	29.82
	DVM (24 özellik) – DVM (110 özellik)	39.65
	RO (24 özellik) – RO (110 özellik)	37.49
Seçilmiş özelliklerin karşılaştırılması	EYK (24 özellik) – DVM (24 özellik)	24.23
	DVM (24 özellik) – RO (24 özellik)	9.05
	RO (24 özellik) – EYK (24 özellik)	17.20

Sınıflandırma algoritmalarının performansları açısından sonuçlar ele alındığında değerlendirmeye alınan tüm veri setleri için DVM ve RO algoritmalarının EYK algoritmasından %2 seviyelerinde daha yüksek sınıflandırma doğruluğuna ulaştığı görülmektedir. Bu durum nesne tabanlı sınıflandırmada makine öğrenme algoritmalarının tematik harita doğruluğuna olumlu etkisinin olduğunu destekler niteliktedir. DVM ve RO algoritmalarının performansı karşılaştırıldığında, DVM’nin tüm veri setleri için en yüksek sınıflandırma doğruluğuna ulaştığı ancak iki algoritmanın genel doğruluğu esas alındığında benzer sınıflandırma sonuçları elde ettiği görülmektedir. Sınıflandırma algoritmalarının en yüksek doğruluğa ulaştığı bilgi kazancı tarafından belirlenen veri seti için sınıflandırma doğrulukları arasındaki farklılıklar Wilcoxon işaretli sıralar testi kullanılarak analiz edildiğinde EYK, DVM ve RO algoritmalarının birbirlerinden istatistiksel olarak farklı performanslar sergilediği görülmüştür (Tablo 4). Elde edilen bu sonuç DVM algoritmasının 24 özellik içeren veri seti kullanımıyla EYK ve RO algoritmalarından istatistiksel olarak daha yüksek sınıflandırma doğruluğuna ulaştığını göstermektedir. Aynı şekilde 24 özellikten oluşan veri seti için, RO ve EYK algoritmalarıyla ulaşılan sınıflandırma doğrulukları arasındaki farkın istatistiksel olarak anlamlı bir fark olduğu belirlenmiştir.

5.SONUÇLAR

Uzaktan algılama teknolojilerinde ve algılayıcı sistemlerinde yaşanan yenilikler kullanıcıları yeni nesil uydu görüntülerinin analizi noktasında güçlü ve etkili alternatif yöntemlerin geliştirilmesi arayışına itmiştir. Nense tabanlı görüntü analizi özellikle benzer spektral özelliklere sahip heterojen yapılu piksellerden oluşan yüksek çözünürlüklü uydu görüntülerinin sınıflandırılması problemlerinin çözümündeki başarısıyla kullanımı ön plana çıkan bir yaklaşımdır. Nesne tabanlı sınıflandırma yaklaşımının temel avantajı segmentasyon sonucu üretilen nesnelere için birçok özelliğin tanımlanabilmesi ve sınıflandırma işleminde bu özelliklerden yararlanılabilmesidir. Bu sayede girdi verisi olarak kullanılan uydu görüntüsünün multispektral bantları yerine yüzlerce banttan (özellikten) oluşan bir veri setinin sınıflandırılması söz konusu olmaktadır. Bu nedenle yüksek boyutlu veri setleri kullanımında yaşanan Hughes fenomeni diğer bir ifadeyle boyutsallık sorunu, özellik seçimi ve veri seti boyutunun azaltılması konuları nesne tabanlı görüntü analizinde de önemli çalışma konuları olarak ortaya çıkmaktadır.

Bu çalışmada nesne tabanlı görüntü analizinde özellik seçimi ve veri boyutunun azaltılması probleminin çözümüne yönelik olarak filtreleme tabanlı ki-kare, bilgi kazancı ve ReliefF algoritmalarının kullanımı ele alınarak, özellik seçiminin sınıflandırma doğruluğu üzerindeki etkileri incelenmiştir. Özellik seçimi ile elde edilen veri setlerinin sınıflandırılması ve tematik harita üretiminde EYK, DVM ve RO algoritmaları kullanılmış ve sınıflandırma performansları detaylı bir şekilde analiz edilmiştir. Çalışma sonuçları uygulamada değerlendirmeye alınan veri setinin nesne tabanlı yaklaşımla sınıflandırılması probleminde özellik seçimi yardımıyla veri boyutu %78 azaltılmasına karşın sınıflandırma doğruluğunda %4'e varan seviyelerde artış olduğunu ve sınıflandırma doğruluğundaki bu değişimin istatistiksel olarak anlamlı olduğunu göstermektedir. Özellik seçimi sonuçları, bilgi kazancı algoritmasının optimum özellikleri içeren veri seti tespitinde ve veri seti boyutu azaltılmasında diğer algoritmalara göre daha başarılı olduğunu göstermektedir. Nitekim, EYK, DVM ve RO algoritmaları söz konusu veri seti kullanarak en yüksek sınıflandırma doğruluğuna ulaşmıştır. Bu durum bilgi kazancı algoritmasının özellik seçimindeki etkinliğini doğrular niteliktedir. Sınıflandırma algoritmalarının performansı karşılaştırıldığında EYK, DVM ve RO algoritmalarının performanslarının veri seti boyutundaki artışa duyarlı olduğu ve sınıflandırma doğruluğunda istatistiksel olarak anlamlı azalmalar olduğu görülmüştür. Seçilen optimum özelliklere sahip veri seti için DVM ile en yüksek sınıflandırma doğruluğuna ulaşıırken, her üç sınıflandırıcı veri seti için istatistiksel olarak farklı sınıflandırma sonuçları üretmiştir. Elde edilen tüm sonuçlar bir arada değerlendirildiğinde nesne tabanlı görüntü analizinde benzer özelliklere sahip görüntü nesnelерinin ayırt edilmesi noktasında özellik seçiminin sınıflandırma doğruluğuna olumlu etkisinin olduğu ve filtreleme tabanlı bilgi kazancı yaklaşımının özellik seçimi için etkili bir algoritma olduğu sonucuna varılmıştır.

KAYNAKLAR

- Baatz, M., Schäpe, A.,** 2000. Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T., Griesebner, G. (Eds.), *Angewandte Geographische Informationsverarbeitung XII*. Wichmann Verlag, Karlsruhe, Germany.
- Baccianella, S., Esuli, A., Sebastiani, F.** 2014. Feature Selection for Ordinal Text Classification. *Neural Computation*, 26(3), 557-591.
- Belgiu, M., Dragut, L., Strobl, J.,** 2014. Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using WorldView-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 205-215.
- Blaschke, T.,** 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2-16.
- Blum, A. L., Langley, P.,** 1997. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1-2), 245-271.
- Breiman, L.,** 2001. Random forests. *Machine Learning*, 45(1), 5-32.
- Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S., and Malo, J.,** 2011. *Remote Sensing Image Processing, Synthesis Lectures on Image, Video and Multimedia Processing*, Bovik, Al C., Ed. LaPorte: Morgan & Claypool Publishers.
- Çölkesen, İ.,** 2015. *Yüksek çözünürlüklü uydu görüntüleri kullanarak benzer spektral özelliklere sahip doğal nesnelere ayırt edilmesine yönelik bir metodoloji geliştirme*, Doktora Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü.
- Demers, A. M., Banks, S. N., Pasher, J., Duffe, J., Laforest, S.** 2015. A Comparative Analysis of Object-Based and Pixel-Based Classification of RADARSAT-2 C-Band and Optical Satellite Data for Mapping Shoreline Types in the Canadian Arctic. *Canadian Journal of Remote Sensing*, 41(1), 1-19.
- Duro, D. C., Franklin, S. E., Dube, M. G.,** 2012. Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. *International Journal of Remote Sensing*, 33(14), 4502-4526.

- 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), 2547-2561.
- Guyon, I., Elisseeff, A.,** 2003. An introduction to variable and feature selection introduction. *Journal of Machine Learning Research*, 3, 1157–1182.
- Hall, M. A.,** 1998. *Correlation-based feature selection for machine learning*, PhD diss., University of Waikato, Hamilton.
- 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), 725-749.
- 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), 352-359.
- Kavzoglu, T., Mather, P. M.** 2002. The role of feature selection in artificial neural network applications. *International Journal of Remote Sensing*, 23(15), 2919-2937.
- Kira, K., Rendell, L.A.,** 1992. The feature selection problem: Traditional methods and a new algorithm, *Proceedings of the Tenth National Conference on Artificial Intelligence*, (pp. 129-134), USA: California, July 12-16.
- Kohavi, R., John, G. H.,** 1997. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2), 273-324.
- Kononenko, I.,** 1994. Estimating Attributes: Analysis and Extensions of RELIEF. In B. Francesco, R. Luc de (Ed.), *ECML-94. Proceedings of the Seventh European Conference on Machine Learning*, (pp. 145-154). Italy: Catania, April 5-8.
- Kumar, V., Minz, S.,** 2014. Feature selection: A literature review, *Smart Computing Review*, 4(3), 211-229.
- Landgrebe, D.A.** (2003). *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ: John Wiley & Sons.
- Ma, L., Cheng, L., Li, M. C., Liu, Y. X., Ma, X. X.,** 2015. Training set size, scale, and features in Geographic Object-Based Image Analysis of very high resolution unmanned aerial vehicle imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102, 14-27.
- Mingers J.,** 1989. An empirical comparison of pruning methods for decision tree induction, *Machine Learning*, 4(2), 227–243.
- Mountrakis, G., Im, J., Ogole, C.,** 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247-259.
- Mwangi, B., Tian, T. S., Soares, J. C.,** 2014. A review of feature reduction techniques in neuroimaging. *Neuroinformatics*, 12(2), 229-244.
- Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q. H.,** 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, 115(5), 1145-1161.
- Pal, M., Foody, G. M.,** 2010. Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5), 2297-2307.
- Plackett, R. L.,** 1983. Pearson, Karl and the Chi-Squared Test. *International Statistical Review*, 51(1), 59-72.
- Radoux, J., Bogaert, P., Fasbender, D. & Defourny, P.** 2011. Thematic accuracy assessment of geographic object-based image classification. *International Journal of Geographical Information Science*, 25(6), 895-911.
- Quinlan J.R.,** 1993. *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.
- Saeys, Y., Inza, I., Larranaga, P.,** 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19), 2507-2517.
- Vapnik, V. N.,** 1995. *The nature of statistical learning theory*. New York, USA: Springer-Verlag.