

MULTISPEKTRAL UYDU GÖRÜNTÜLERİ İÇİN EN UYGUN BANT SEÇİMİNİN SINIFLANDIRMA DOĞRULUĞUNA ETKİLERİNİN İNCELENMESİ

T. Kavzoğlu *, İ. Çölkesen, E.K. Şahin

Gebze Yüksek Teknoloji Enstitüsü, Jeodezi ve Fotogrametri Müh. Bölümü, 41400 Gebze-Kocaeli
(kavzoglu@gyte.edu.tr, icolkesen@gyte.edu.tr, e.sahin@gyte.edu.tr)

ANAHTAR SÖZCÜKLER: Arazi kullanımı, Özellik seçimi, Karar ağacı, Naive bayes, Mutispektral veri, McNemar's test

ÖZET:

Uydu görüntüleri yardımıyla arazi örtüsü veya kullanım haritalarının üretilmesi uzaktan algılama teknolojilerinin yaygın olarak kullanıldığı uygulama alanlarından birisidir. Uzaktan algılama uydularının sahip olduğu sensörler farklı konumsal ve spektral çözünürlüklerde yeryüzündeki objeler hakkında bilgi toplayabilme kabiliyetine sahiptirler. Uydular üzerindeki mevcut sensörlerin spektral bant sayıları sınırlı sayıda olabileceği gibi (Spot HRV, Landsat MSS ve TM) çok sayıda banta da (Chris Proba, Hyperion CASI ve AVIRIS) sahip olabilmektedir. Spektral bantların her birisi kendi başına uzaktan algılanan objelerin tespiti açısından önemli bilgiler içermesine karşın veri setlerinin boyutunun artırılmasıyla bantlar arasında yüksek kolerasyon bir problem ortaya çıkmaktadır. Söz konusu problemin ortadan kaldırılması veya etkisinin azaltılması için kolerasyonlu bantların elemine edilerek en uygun bantların seçilmesi yoluna gidilmektedir. Bu çalışmada, Gebze ilçesine ait genel arazi örtüsü/kullanımını temsil eden tematik harita üretiminde 2002 tarihinde kaydedilen Landsat TM, EO-ALI ve Terra ASTER multisensör ve multitemporal veri setlerinden yararlanılmıştır. Görüntülerin birleştirilmesi ile elde edilen 24 bantlı veri seti içerisinde en uygun bantların tespit edilmesinde, geleneksel ayrılabilirlik indeksleri yerine karar ağaçları, en yakın komşuluk algoritması ve naive bayes metodları kullanılmıştır. Bant kombinasyonları için elde edilen performansların değerlendirilmesinde genel sınıflandırma doğrulukları dikkate alınmıştır. Bu doğruluklar arasındaki farkların istatistiksel olarak anlamlılığı McNemar's istatistiksel testi kullanılarak analiz edilmiştir. Çalışma sonucunda, 9 bant sayısına kadar sınıflandırma doğruluğunda istatistiksel olarak anlamlı değişimler olduğu görülmüştür. Diğer taraftan, 9 banttan sonra bant sayısındaki artışın sınıflandırma doğruluğuna etkisinin istatistiksel olarak anlamlı olmadığı, dolayısıyla bu noktadan sonra elde edilen performansların aynı olduğu belirlenmiştir.

1. GİRİŞ

Uzaktan algılama teknolojileri belirli bir zamanda belirli bir çözünürlükte yeryüzü üzerindeki objeleri görüntüleyebilme kabiliyetine sahiptir. Uzaktan algılama da kullanılan uydular sahip oldukları sensör tipine göre farklı konumsal ve spektral çözünürlüklerde yeryüzünü görüntülemektedir. Sensörlere ait spektral kanalların her birisi kendi başına uzaktan algılanan objelerin tespiti açısından önemli bilgiler içermektedir. Buna karşın spektral çözünürlük arttıkça bantlar arasında yüksek kolerasyon bir problem olarak ortaya çıkmaktadır. Bu durum çok bantlı veri setlerinde bilgi fazlalığını ortaya çıkarmaktadır. Bu problem Hughes etkisi veya boyutsallık problemi (curse of dimensionality) olarak bilinmekte ve ilk yüksek spektral çözünürlüklü verilerin elde edildiği 1980'li yılların sonundan beri uzaktan algılama uygulamalarında karşılaşılmaktadır (Bellman,1961; Hughes, 1968; Landgrebe, 2003). Sınıflandırma metodlarının etkinliğinin artırılması için söz konusu bilgi fazlalığının veya yüksek kolerasyona sahip bantların elemine edilmesine ihtiyaç duyulmaktadır. Bu nedenle çok sayıda spektral banta sahip uzaktan algılanmış görüntülerin sınıflandırılması probleminin çözümünde en uygun bantların tespit edilmesi önemli bir ihtiyaçtır. Tüm veri seti içerisinde birtakım değerlendirme ölçütleri göz önüne alınarak en uygun alt kümenin aranması işlemi özellik seçimi olarak adlandırılmaktadır. Özellik seçiminde temel amaç fazlalık veya kolerasyonlu bantların elemine edilmesidir. Tüm veri seti içerisindeki tüm olası altkümelerin belirlenmesi büyük bir hesaplama yükü gerektirdiğinden, günümüze kadar birçok tarama tekniği geliştirilmiştir (genetik algoritma, bulanık mantık vb.).

En uygun dalga boyu aralığının belirlenmesi, kullanılan bant sayısı, konumsal ve spektral çözünürlük kavramları yüksek spektral çözünürlüklü görüntülerin analizinde karşılaşılan en önemli problemlerdendir. Bu görüntüler birçok uygulama için önemli bilgiler içermektedir. Buna karşın bazı özel uygulamalar için söz konusu bantların hepsinin kullanılmasına gerek yoktur. Diğer bir ifadeyle bazı bantlar önemli iken bazı bantlardaki bilgiye ihtiyaç duyulmamaktadır (Jia et al., 2013). Herhangi bir bant belirli bir problemin çözümünde kullanılabilirken, diğer bir problem için kullanılamayabilir ve farklı uygulama alanları için her bir bantın sahip olduğu dalga boyu aralığı birer aday niteliğindedir.

Veri setini temsil eden sınıf karakteristiklerinin modellenmesinde en uygun bantların belirlenmesi sınıflandırıcının performansını etkileyen önemli bir unsurdur. Herhangi bir sınıflandırma metodu girdi değişkenler/özelliklerin iyi kalitede olduğu ve yeterli seviye de temsil edildiğinde en iyi sonucu verir. Genelde veri madenciliğinde tüm spektral bantların kullanılması iyi bir çözüm değildir. Gereksiz hesap yükünün yanında, yüksek korelasyona sahip çok boyutlu bir veri seti ile işlem yapılması söz konusu olmaktadır. Sınıflandırmada kullanılan eğitim örneklerinin sayısının kullanılan bant sayısına oranı düşük olduğunda düşük genelleştirme kapasitesine sahip olur. Bu durum, sınırlı sayıdaki eğitim örnekleri ile elde edilen sınıflandırma doğruluğunun yüksek, test veri seti için doğruluğun düşük olmasına neden olur. En etkili bantların belirlenmesi genel olarak çok bantlı veri seti içerisinde en az sayıda bant ile en doğru sonuçların elde edilmesidir. Kuo and Landgrebe (2004) özellik seçimi ve özellik çıkarımının önemli bir ön işlem olduğunu vurgulamışlardır. Bu ön işlem adımının gerçekleştirilmesi ilk

uygulamalarda tüm aday bantlar içerisinde en uygun bant kombinasyonunun belirlenmesi prensibine dayanmaktaydı. Daha sonra dönüşüm matrisleri kullanarak çeşitli dönüşümlerle daha az sayıda banttan oluşan yeni bir veri setinin oluşturulması yoluna gidilmiştir. Son yıllarda sadece spektral özelliklerin kullanılmasındaki kısıtlamalar fark edilerek, içeriksel bilginin de değerlendirilmesi ön plana çıkmıştır (De Martino et al., 2003; Jia, 2006; Chen et al., 2011).

Bu çalışmada, 2002 tarihli Landsat TM, EO-ALI ve Terra ASTER uydu görüntülerinin birleştirilmesi ile elde edilen 24 banta sahip multispektral veri seti ile Gebze ilçesinin genel arazi kullanımı ve arazi örtüsünün temsil eden tematik haritanın üretilmesi hedeflenmiştir. Çalışmanın temel amacı veri seti içerisinde istenilen bant sayısı dikkate alınarak tüm olası kombinasyonlar içerisinde en yüksek sınıflandırma doğruluğunu veren bant kombinasyonunun tespit edilmesidir. En yüksek sınıflandırma doğruluğunu veren bant kombinasyonunun tespit edilmesinde karar ağaçları, en yakın komşuluk ve naive bayes metotları özellik seçimi metodu olarak kullanılmıştır. Elde edilen sınıflandırma doğrulukları arasındaki farkların istatistiksel olarak anlamlılığı McNemar's testi kullanılarak analiz edilmiştir.

2. ÇALIŞMA ALANI VE KULLANILAN VERİ

Marmara bölgesinin en büyük ikinci ilçesi olan Gebze, Kocaeli endüstrisinin büyük bir bölümünü içerisinde barındırmaktadır. Yaklaşık yüz ölçümü 500 km² olan ilçe, Kocaeli ilinin kuzeybatısında, İstanbul'un doğusunda yer almaktadır (Şekil 1). 2012 adrese dayalı nüfus kayıt sistemi verilerine göre ilçenin toplam nüfusu 402.650 kişidir (www.tuik.gov.tr).



Şekil 1. Çalışma alanı (Gebze) ilçe sınırları.

Bu çalışmada Gebze ilçesinin genel arazi kullanımı ve arazi örtüsünün belirlenmesi amacıyla 2002 tarihinde kaydedilen Landsat TM, EO-ALI ve Terra ASTER uydu görüntüleri kullanılmıştır. Uydu görüntüleri bölgeye ait 1/25.000 ölçekli haritalardan yararlanılarak geometrik olarak düzeltilerek UTM projeksiyon sistemine dönüşümü gerçekleştirilmiştir.

3. ÖZELLİK SEÇİMİ İÇİN KULLANILAN ALGORİTMALAR

3.1. Karar Ağaçları Algoritması

Karar ağaçları son yıllarda literatürde sıkça kullanılan sınıflandırma ve örüntü tanımlama algoritmalarından birisidir. Bu yöntemin yaygın olarak kullanımının en önemli nedeni ağaç

yapılarının oluşturulmasında kullanılan kuralların anlaşılabilir ve sade olmasıdır. Karar ağaçları (KA) arazi örtüsünün sınıflandırılmasında çok aşamalı veya ardışık bir yaklaşım kullanılmaktadır. Yöntem kompleks yapıdaki bir sınıflandırma problemini çok aşamalı bir hale getirerek basit bir karar verme işlemi gerçekleştirir (Safavian and Landgrebe, 1991; Kavzoğlu ve Çölkese, 2010). Bir karar ağacının temel yapısı düğüm, dal ve yaprak olarak adlandırılan üç temel kısımdan oluşur. Bu ağaç yapısında her bir öznitelik bir düğüm tarafından temsil edilir. Dallar ve yapraklar ağaç yapısının diğer elemanlarıdır. Ağaçta en son kısım yaprak en üst kısım ise kök olarak adlandırılır. Kök ve yapraklar arasında kalan kısımlar ise dal olarak ifade edilir (Quinlan, 1993). Eğitim verileri ve bu verilere ait öznitelik bilgilerinden yararlanılarak bir karar ağacı yapısı oluşumunda temel prensip verilere ilişkin bir dizi sorular sorulması ve elde edilen cevaplar doğrultusunda karar kuralları oluşturulmasıdır. Ağacın ilk düğümü olan kök düğümünde verilerin sınıflandırılması ve ağaç yapısının oluşturulması için sorular sorulmaya başlanır ve dalları olmayan düğümler ya da yapraklara gelene kadar bu işlem devam eder (Pal and Mather, 2003).

3.2. En Yakın Komşuluk Algoritması

Bu yöntem, sınıfları belli olan bir örnek kümesindeki piksel değerlerinden yararlanarak, örneğe katılacak yeni bir pikselin hangi sınıfa ait olduğunu belirlemek amacıyla kullanılır. Bu algoritmalar eğitim verisinden en çok bezeyen tek bir örneği bulmak için uzaklık fonksiyonları kullanırlar. Bulunan örnek yeni bir örneğin sınıflandırılması için kullanılır (Hart, 1968; Gates, 1972). En yakın komşuluk algoritması öznitelik değeri bilinmeyen bir piksele en yakın eğitim pikselinin bulunmasında Öklit uzaklığını kullanmaktadır. Bu durumda test pikseli için öznitelik değeri belirlenen eğitim pikseline ait öznitelik değeri olarak atanır.

3.3. Naive Bayes Algoritması

Olasılık teoremine dayalı Naive Bayes sınıflandırıcısı matematiksel olarak anlaşılabilir ve literatürde sık kullanılan kontrollü öğrenme algoritmasıdır (John and Langley, 1995; Domingos and Pazzani, 1997). Literatürde, algoritmanın basit yapısına rağmen karmaşık sınıflandırma problemlerinin çözümünde iyi performans sergilediği ifade edilmektedir. Ayrıca bu algoritmanın en önemli özelliği kullanılan veri seti içerisindeki az miktardaki gürültünün sonuç üzerindeki etkilerinin az olmasıdır. Naive Bayes algoritması Bayes teoremine dayalıdır ve aşağıdaki temel eşitlikle ifade edilebilir (Rish, 2001).

$$P(X|C) = \prod_{i=1}^n P(X_i|C) \quad (1)$$

Bu eşitlikte $X = (X_1, \dots, X_n)$ özellik vektörünü, C ise sınıfı göstermektedir.

4. UYGULAMA

Çalışma alanında yapılan ön çalışmalar neticesinde Gebze ilçesinin arazi örtüsü ve kullanımını temsil eden altı temel sınıfın (su, iğne yapraklı orman, geniş yapraklı orman, bozkır, toprak-taş ve şehir) mevcut olduğu tespit edilmiştir. Sınıflandırmada ve en uygun bant kombinasyonunun belirlenmesinde kullanılmak üzere örnekleme alanları mevcut haritalar, hava fotoğrafları ve araziden el-GPS'i ile toplanan veriler yardımıyla oluşturulmuştur. Belirlenen örnekleme

alanları içerisinde rastgele örnekleme prensibi esas alınarak eğitim ve test veri setleri oluşturulmuştur. Hesaplanan genel sınıflandırma doğruluklarının karşılaştırılmasında objektifliğin sağlanması amacıyla eğitim ve test veri setleri belirlenirken eşit sayıda örnek seçilmiştir. Sonuç olarak belirlenen altı arazi örtüsü sınıfı için eğitim verisi olarak 2.520 piksel (her bir sınıf için 420 piksel), test verisi olarak toplam 1.980 piksel (her bir sınıf için 330 piksel) seçilmiştir.

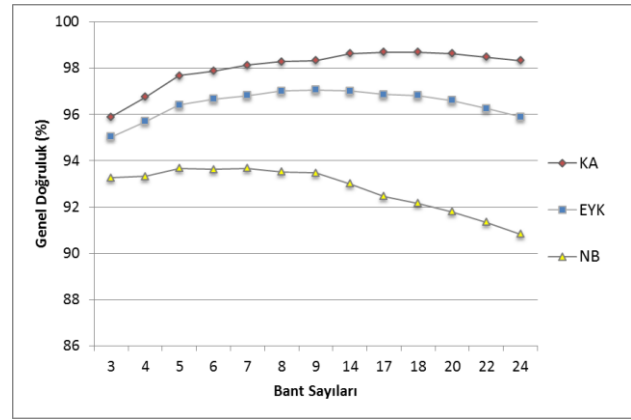
Çalışmada 24 bantlı multispektral veri seti içerisinde en uygun 3, 4, 5, 6, 7, 8, 9, 14, 17, 18, 20 ve 22 bant kombinasyonlarını tespit etmek için öncelikle söz konusu bant sayıları için olası tüm kombinasyonlar belirlenmiştir. Bant sayısındaki artışla birlikte yöntemlerin sınıflandırma performansındaki değişimlerin belirlenmesinde öncelikli olarak sınıflandırma sonucunda elde edilen genel doğruluklar kullanılmıştır. Bu değerlendirmenin yanı sıra hesaplanan genel doğruluklar arasındaki farklılıkların anlamlılığı McNemar's testi kullanılarak analiz edilmiştir. McNemar's testi, iki farklı sınıflandırma yöntemi veya iki farklı eğitim seti için elde edilen genel sınıflandırma doğruluğu arasındaki farkın istatistiksel olarak anlamlı olup olmadığının tespitinde kullanılan istatistiksel bir testtir. McNemar's testi, sınıflandırma sonucu test veri seti için tahmin edilen sınıf etiketleri ile orijinal sınıf etiketlerinin karşılaştırılması esasına dayanmaktadır (Foody, 2004). Hesaplanan test değeri %95 güven aralığındaki $h=3.84$ kritik değerinden büyük olduğunda iki sınıflandırma sonucu arasındaki farkın istatistiksel olarak anlamlı olduğunu, dolayısıyla iki sınıflandırıcıya ait performansların birbirinden farklı olduğunu ifade edebiliriz.

Belirlenen 12 farklı bant sayısı için ayrı ayrı tüm kombinasyonlar belirlenmiş ve oluşturulan eğitim veri setleri kullanılarak karar ağaçları (KA), en yakın komşuluk (EYK) ve naive bayes (NB) yöntemleri ile sınıflandırmaya esas olacak modeller oluşturulmuştur. Elde edilen modeller test veri setleri kullanılarak analiz edilmiştir. Örneğin 24 bantlı multispektral veri seti içerisinde en uygun 3 bantın seçilmesi probleminde, veri setinin 3 bantlı 2024 kombinasyonu mevcuttur. Söz konusu kombinasyonlar içerisinde en yüksek sınıflandırma doğruluğunu veren kombinasyonun tespitinde KA, EYK ve NB metodları kullanılmıştır.

Tablo 1. En uygun bant kombinasyonları için elde edilen sınıflandırma sonuçları.

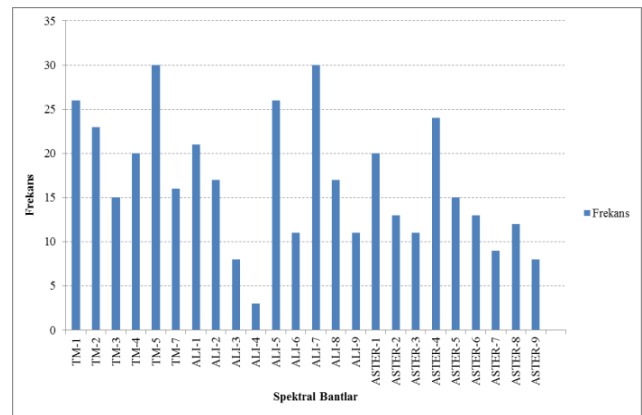
Bant Kombinasyonu	KA (%)	EYK (%)	NB (%)
3	95,901	95,040	93,269
4	96,761	95,698	93,320
5	97,672	96,407	93,674
6	97,874	96,660	93,623
7	98,128	96,812	93,674
8	98,279	97,014	93,522
9	98,330	97,065	93,472
14	98,634	97,014	93,016
17	98,684	96,862	92,460
18	98,684	96,812	92,156
20	98,634	96,609	91,802
22	98,482	96,255	91,346

Bu prensipten hareketle 12 farklı bant sayısı için yöntemler tarafından test veri seti için hesaplanan en yüksek sınıflandırma doğrulukları Tablo 1'de verilmiştir. Çalışmada kullanılan üç sınıflandırma metodu için elde edilen sınıflandırma doğrulukları karşılaştırıldığında en yüksek sınıflandırma doğruluğunun KA sınıflandırıcısı ile elde edildiği görülmektedir. Ayrıca düşük sayıda özellik seçiminde KA ve EYK yöntemlerinin benzer sonuçlar verdiği, her iki yönteminde NB sınıflandırıcısına göre yaklaşık %3 daha yüksek doğruluklu sonuçlar ürettiği görülmüştür. Bant sayısındaki artışla birlikte tüm sınıflandırma yöntemleri için hesaplanan sınıflandırma doğrulukları 9 banta kadar artış gösterirken, bu noktadan sonra KA ve EYK yöntemlerinin genel doğruluklarında önemli bir değişiklik olmadığı NB yönteminin sınıflandırma performansının ise azalma eğiliminde olduğu görülmektedir (Şekil 2).



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

KA, EYK ve NB yöntemleri ile elde edilen en uygun bant kombinasyonları içerisinde multispektral veri setinin hangi bandının ne kadar sıklıkta kullanıldığını belirlemek amacıyla Şekil 3'te gösterilen frekans dağılımı hesaplanmıştır. Şekilden de görüleceği üzere Landsat TM görüntüsünün 5., 1., 2. ve 4. bantlarının, EO-ALI görüntüsünün 7., 5. ve 1. bantlarının ve Terra ASTER görüntüsünün 4., 1. ve 5. bantlarına ait frekans değerlerinin yüksek olduğu görülmektedir. Diğer taraftan multispektral veri setinin tüm bantları içerisinde EO-ALI görüntüsünün 4. bandı için en düşük frekans değeri hesaplanmıştır.



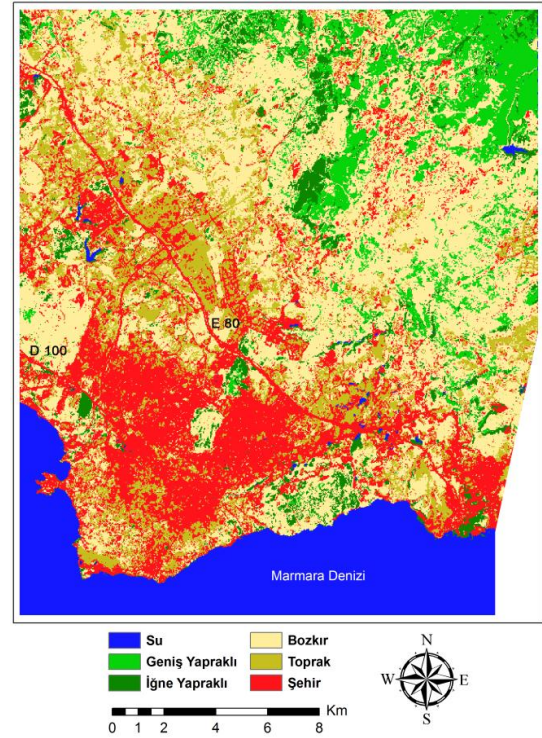
Şekil 3. En uygun bant kombinasyonları için hesaplanan frekans dağılımı.

Yöntemlerin sınıflandırma performansları arasındaki farklılıkların istatistiksel olarak anlamlılığı McNemar's testi ile analiz edilmiştir. Bu amaçla tüm bant kombinasyonları için, test veri seti için hesaplanan sınıf etiketleri ile orijinal sınıf etiketleri kullanılarak hesaplanan istatistik değerleri Tablo 2'de gösterilmiştir. Tablodan da görüleceği üzere, KA ile EYK yöntemleri, KA ve NB yöntemleri ve EYK ve NB yöntemleri için yapılan istatistik testi sonuçları $h=3.84$ kritik değerinden büyüktür. Yalnızca KA ve EYK yöntemlerinin karşılaştırılmasında 3 bant kombinasyonu için hesaplanan test sonucu kritik değerden küçük olduğu görülmektedir. Test sonuçları KA yöntemiyle EYK yönteminin sınıflandırma performansları arasındaki farkın sadece 3 bant kombinasyonunda istatistiksel olarak anlamsız olduğu, diğer tüm durumlar için sınıflandırma doğrulukları arasındaki farkların istatistiksel olarak anlamlı olduğu ifade edilebilir. Elde edilen bu sonuç çalışmada dikkate alınan 12 farklı bant sayısı için KA yönteminin EYK ve NB yöntemlerine göre daha yüksek doğrulukta sınıflandırma sonuçları ürettiğini destekler niteliktedir. Bununla birlikte EYK ve NB arasında hesaplanan istatistiksel test sonuçlarının tamamı kritik değerden büyük olup, tüm durumlarda EYK ve NB arasındaki sınıflandırma performanslarının birbirinden farklı olduğunu göstermektedir.

Tablo 2. Yöntemlerinin performans analizinde hesaplanan McNemar's test değerleri.

Bant Kombinasyonu	KA - EYK	KA - NB	EYK - NB
3	2.813	9.714	24.538
4	4.598	18.400	41.565
5	7.890	26.009	60.238
6	32.513	8.430	68.890
7	8.929	33.223	71.406
8	9.143	40.920	74.560
9	11.294	44.144	75.208
14	17.161	47.163	92.366
17	23.558	56.458	107.079
18	25.412	60.890	112.993
20	26.224	62.667	118.914
22	27.191	59.458	126.452

En iyi bant kombinasyonları kullanılarak çalışma alanının genel arazi örtüsünü gösteren tematik haritalar üretilmiştir. Söz konusu tematik haritalara örnek olarak KA sınıflandırıcısıyla belirlenen 9 bantlı en iyi kombinasyonla oluşturulan harita Şekil 4'de gösterilmiştir. Tematik harita incelendiğinde çalışma alanının %20'lik kısmını kapsayan şehirleşmenin özellikle kıyı bölgelerinde ve ana karayolları (D-100 ve E-80) boyunca yoğunlaştığı görülmektedir. Çalışma alanındaki bitki örtüsü dağılımı analiz edildiğinde toplam alanın %35'lik kısmını bozkır/mera alanlarının kapladığı, geniş yapraklı orman alanlarının ise özellikle çalışma alanının Karadeniz'e kıyısı buluna Kuzey-Doğu kesiminde yoğunlaştığı görülmektedir.



Şekil 4. Karar ağaçları yöntemiyle belirlenen en iyi 9 bant kombinasyonu ile üretilen tematik harita.

5. SONUÇLAR

Günümüzde uzaktan algılama teknolojilerinde yaşanan gelişmeler neticesinde yüksek spektral ve konumsal çözünürlüklere sahip görüntüler elde edilebilmektedir. Yüksek spektral çözünürlüklere sahip görüntüler yeryüzü ile ilgili zengin bir bilgi kaynağı sağlamasına karşın, söz konusu görüntülerin işlenmesinde birtakım zorluklar vardır. Bunların başında bilgi fazlalığı, birbirleriyle yüksek kolarezyona sahip ve gürültü içeren veriler ve hesaplama için gerekli süre gelmektedir. Yüksek spektral çözünürlüğe sahip verilerin analizinde karşılaşılan en önemli problem en uygun dalga boylarının ve kullanılacak optimum bant sayısının belirlenmesidir.

Bu çalışmada 24 bantlı multispektral veri seti yardımıyla Gebze ilçesine ait arazi örtüsü haritasının üretilmesi problemi ele alınmıştır. Yüksek spektral çözünürlüğe sahip bu veri seti içerisinde en uygun bantların seçilmesinde, geleneksel ayrılabilirlik indeksleri yerine karar ağaçları (KA), en yakın komşuluk (EYK) algoritması ve naive bayes (NB) metotları kullanılmıştır. Çalışmanın temel amacı veri seti içerisinde istenilen bant sayısı dikkate alınarak tüm olası kombinasyonların belirlenmesi ve bu kombinasyonlar içerisinde en yüksek sınıflandırma doğruluğu veren bant kombinasyonunun tespit edilmesidir.

Çalışmada kullanılan yöntemlerin sınıflandırma performansları analiz edildiğinde KA ve EYK yöntemlerinin NB yöntemine göre %8'lere varan oranda daha iyi sonuçlar verdiği görülmüştür. Bant sayısındaki artışa karşılık sınıflandırma performanslarındaki değişimler incelendiğinde özellikle çok bantlı veri seti kullanıldığı durumlarda NB yönteminin performansında önemli derecede bir düşüş olduğu gözlemlenmiştir. Diğer taraftan KA ve EYK yöntemlerinin performansları kritik bant sayısı olan 9 banttan sonraki artışlarda benzer sınıflandırma performansları gösterdiği

belirlenmiştir. Çalışma sonucunda, 9 banta kadar bant sayısındaki artışa paralel olarak sınıflandırma doğruluğunda istatistiksel olarak anlamlı değişimler olduğu görülmüştür. Diğer taraftan, 9 banttan sonra bant sayısındaki artışın sınıflandırma doğruluğuna etkisinin istatistiksel olarak anlamlı olmadığı, dolayısıyla bu noktadan sonra elde edilen performansların aynı olduğu ifade edilebilir. Diğer bir ifadeyle, en uygun 9 bant ile üretilen tematik harita doğruluğunun, 14, 17, 18, 20, 22 ve 24 bant kullanılarak üretilen tematik haritalar ile istatistiksel olarak aynı olduğu ifade edilebilir.

KAYNAKLAR

Bellman, R.E.,1961. *Adaptive Control Processes*. A Guided Tour. Princeton, NJ: Princeton University Press.

Chen, X., Fang, T., Huo, H., Li, D.R., 2011. Graph-Based Feature Selection for Object-Oriented Classification in VHR Airborne Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1), pp. 353-365.

De Martino, M., Causa, F., Serpico, S.B.,2003. Classification of optical high resolution images in urban environment using spectral and textural information. *IEEE International Geoscience and Remote Sensing Symposium*, 1, pp. 467-469.

Domingos P., Pazzani, M.,1997. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2-3), pp. 103-130.

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.

Gates, G.W., 1972. The Reduced Nearest Neighbor Rule. *IEEE Transactions on Information Theory*, 18(3), pp. 431-433.

Hart, P.E., 1968. The Condensed Nearest Neighbor Rule. *IEEE Transactions on Information Theory*, 14(3), pp. 515-516.

Hughes, G.F., 1968. On the mean accuracy of statistical pattern recognition. *IEEE Transactions on Information Theory*, 14(1), pp. 55-63.

Jia, X., 2006. Class modeling and remote sensing image classification using selected spectral and spatial features. *International Congress of Imaging Science*, Rochester, NY, May 7-11.

Jia, X.P., Kuo, B.C., Crawford, M.M, 2013. Feature Mining for Hyperspectral Image Classification. *Proceedings of the IEEE*, 101(3), pp. 676-697.

John, G.H., Langley, P., 1995. Estimating Continuous Distributions in Bayesian Classifiers. *Eleventh Conference on Uncertainty in Artificial Intelligence*, San Mateo, pp. 338-345.

Kavzoğlu, T., Çölkesen, İ., 2010. Karar ağaçları ile uydu görüntülerinin sınıflandırılması: Kocaeli örneği. *Harita Teknolojileri Elektronik Dergisi*, 2(1), pp. 36-45.

Kuo, B.C., Landgrebe, D.A., 2004. Nonparametric weighted feature extraction for classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42(5), pp. 1096-1105.

Landgrebe, D.A.,2003. *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ:Wiley.

Pal M., Mather P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4), pp. 554-565.

Quinlan J.R., 1993. *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA, 302 p.

Rish, I., 2001. An empirical study of the Naïve Bayes classifier. International Joint Conferences on Artificial Intelligence. 4 - 10 August, Seattle, Washington, USA.

Safavian S.R., Landgrebe D., 1991. A survey of decision tree classifier methodology, *IEEE Transactions on Systems Man and Cybernetics*, 21(3), pp. 660-674.