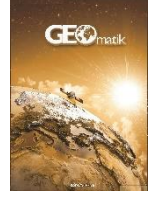




GEOMATİK

<https://dergipark.org.tr/tr/pub/geomatik>

e-ISSN 2564-6761



Şekil göstergeleri ve topluluk öğrenmesi sınıflandırma algoritmaları ile bina detaylarının şekil karmaşıklık analizi

Hüseyin Safa Duman*¹, Melih Başaraner²

¹Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Harita Mühendisliği Anabilim Dalı, UA ve CBS Programı, İstanbul, Türkiye

²Yıldız Teknik Üniversitesi, İnşaat Fakültesi, Harita Mühendisliği Bölümü, Kartografya Anabilim Dalı, İstanbul, Türkiye

Anahtar Kelimeler

Şekil Karmaşıklığı
Bina Detayları
Şekil Göstergeleri
Rastgele Orman
Gradyan Artırma

ÖZ

Şekil analizi, bilgisayar görüşü, coğrafi bilgi bilimi, kartografya, uzaktan algılama, kent morfolojisi, arazi yönetimi ve ekoloji gibi çeşitli alanlarda mekansal olguları/nesneleri karakterize etmek ve mekansal örüntüleri ortaya çıkartmak için kullanılır. Bu bağlamda, şekil göstergeleri, genel olarak mekansal detayların geometrilerinin ve/veya onlardan türetilen yardımcı geometrilerin metrik özellikleri yardımıyla karmaşıklık ve benzerlik gibi şekilsel karakteristikleri niceliksel olarak ifade ederler. Bununla birlikte, şekil göstergeleri mekansal detayların farklı şekilsel özelliklerini ölçmektedir. Bu nedenle, bir detayı şekilsel olarak karakterize ederken tek bir şekil göstergesinin kullanımı her zaman yeterli olmaz. Ayrıca, bu amaçla uygun sınıflandırma yöntemlerinin kullanılması da önemlidir. Bu çalışmada, dairesellik, dışbükeylik ve dikdörtgensellik şekil göstergeleri ile rastgele orman ve gradyan artırma topluluk öğrenme sınıflandırma algoritmaları birlikte kullanılarak 300 adet bina detayı şekilsel karmaşıklık düzeylerine göre basit, orta ve karmaşık olarak sınıflandırılmıştır. Görsel algıya dayalı olarak etiketlenen veri setiyle karşılaştırıldığında rastgele orman algoritması %93.33 genel doğruluk ($\kappa = 0.900$) üretirken, gradyan artırma algoritması ise %92.33 genel doğruluk ($\kappa = 0.885$) üretmiştir. Bu bulgular, bina detaylarının şekilsel karmaşıklık düzeylerinin, çeşitli şekil göstergeleri ve yaygın kullanılan topluluk öğrenmesi sınıflandırma algoritmaları aracılığıyla oldukça yüksek bir doğrulukla sınıflandırılabileceğini göstermiştir.

Shape complexity analysis of building features with shape indices and ensemble learning classification algorithms

Keywords

Shape Complexity
Building Features
Shape Indices
Random Forest
Gradient Boosting

ABSTRACT

Shape analysis is used to characterize spatial phenomena/objects and reveal spatial patterns in various fields such as computer vision, geographic information science, cartography, remote sensing, urban morphology, land management and ecology. In this context, shape indices quantitatively describe the morphological properties such as complexity and similarity usually through the metric properties of the geometries of spatial features and/or the auxiliary geometries derived from them. However, shape indices measure different shape characteristics of spatial features. Therefore, the use of a single shape index is not always sufficient when characterizing a feature morphologically. In addition, it is also important to use appropriate classification methods for this purpose. In this study, using circularity, convexity and rectangularity shape indices together with random forest and gradient boosted ensemble learning algorithms, 300 building features were classified as simple, moderate and complex by their shape complexity. Compared to the benchmark data generated based on visual perception, the random forest algorithm produced 93.33% overall accuracy ($\kappa = 0.900$) while the gradient boosting algorithm produced 92.33% overall accuracy ($\kappa = 0.885$). These findings showed that the shape complexity levels of building features can be categorized with quite high accuracy if various shape indices are used in conjunction with widely used ensemble learning classification algorithms.

1. Giriş

Şekil, mekansal olguların karakteristiklerinin ve örüntülerinin incelenmesinde kullanılan bir bileşen olup, mekansal analizde önemli bir rol oynar. Şekil tanıma yöntemleri, mekansal veriler ile çalışan bilgisayar görüşü, kent morfolojisi, coğrafya, kartografya, arazi yönetimi ve uzaktan algılama gibi birçok alanda kullanılmaktadır. Şekil analizi, mekansal olguların karşılaştırılması, eşleştirilmesi ve gruplandırılması gibi işlemleri kapsar (Ai ve Ark., 2013; Basaraner ve Cetinkaya, 2017; Da Costa ve Cesar, 2009; Wentz, 1997).

Bilgisayar görüşü kapsamında şekil tanıma yöntemleri ve analizine oldukça yaygın olarak başvurulmaktadır. Tıp alanında tümörlerin tanımlanması, sınıflandırılması, anatomik yapıların değişimleri ve deformasyonlarının ölçülmesi gibi çalışmalara katkı sağlamaktadır. Hücrelerin şekli ve işlevleri arasındaki ilişkinin belirlenmesine dayalı olarak tanımlama/sınıflandırılması (örn. beyaz kan hücreleri) ve ayrıca büyüme ve şekil değişikliklerinin tespiti gibi amaçlarla kullanılmaktadır. Güvenlik alanında yüz, parmak izi, iris tanıma, insan yürüyüşünün analizi ve imza doğrulama gibi uygulamalara hizmet etmektedir (Da Costa ve Cesar, 2009).

Mekansal detayların şekilsel analizi, genelleştirme, geometrik veri kalitesi incelemesi, mekansal veri tabanı güncelleme, kent simgesi (referans nesne) saptaması, semantik sınıflandırma, detay eşleştirme, kent yayılımının incelenmesi ve uydu görüntülerinden otomatik detay çıkartımı gibi çeşitli uygulamalarda kullanılmaktadır. Genelleştirme, detayların tekli olarak genelleştirilmesi (bağımsız genelleştirme) ve detay gruplarının genelleştirilmesi (bağlamsal genelleştirme) adımlarından meydana gelir. Bağımsız genelleştirme aşamasında, detayların şekilsel özelliklerine göre uygulanacak genelleştirme işlemleri ve/veya parametreleri farklılık gösterebilir. Ayrıca, bağımsız genelleştirme sonuçlarının değerlendirilmesinde şekil analizinden yararlanılabilir. Bağlamsal genelleştirme ise yakınlık, büyüklük, şekil ve semantik özelliklerine göre birbiriyle ilintili detayların otomatik olarak saptanmasını ve grup halinde genelleştirilmesini gerektirir. Bu kapsamda, benzer veya aynı şekle sahip detayların belirlenmesi ve buna uygun genelleştirme işleminin otomatik seçimi önemlidir (Basaraner ve Cetinkaya, 2017; Başaraner, 2005; Cetinkaya, 2014). Ayrıca, OpenStreetMap (OSM) gibi gönüllü coğrafi bilgi (volunteered geographic information) ve katılımcı harita yapımı (collaborative mapping) uygulamalarında kitle kaynaklı verilerin resmi verilere kıyasla geometrik kalitesini - özellikle ayrıntı düzeyi anlamında - değerlendirmek için kullanılabilir (Basaraner, 2020). Mekansal veri setlerindeki değişikliklerin saptanması veya veri tabanı içeriklerinin daha kolay ve işlevsel bir biçimde güncellenmesine katkı sağlar (Zhou ve ark., 2018). Ayrıca, kent simgesi/referans detay (landmark) olarak adlandırılan ve özellikle insan ya da araç navigasyonu için önem arz eden dikkat çekici detayların saptanmasında kullanılabilir (Grabler ve Ark., 2008). Benzer geometrik ve topolojik özelliklere sahip detayların aynı sınıfa ait olma olasılığı daha yüksek olduğu için şekil analizi, semantik özellikleri bilinmeyen

detayların sınıflarının belirlenmesine yardımcı olabilir (Huang ve Ark., 2013). Farklı veri setleri bütünleştirilirken, eşlenik detayların belirlenmesinde alansal örtüşme oranına ilave olarak şekil analizinden yararlanılabilir. Kentsel yayılma ve büyüme dinamiklerinin belirlenmesinde ve arazi kullanımının sınıflandırılmasında da şekil analizinden faydalanılmaktadır (Frenkel ve Ashkenazi, 2008; Li ve ark., 2013; Wentz, 1997; Zhong ve Ark., 2020). Arazi toplulaştırmasında düzenli şekillere sahip yeni parsellerin oluşturulması amacıyla şekil analizinden yararlanılabilir (Demetriou ve ark., 2013). Ayrıca, uzaktan algılama ile elde edilen görüntülerden detay çıkartımı işlemine katkı sağlayabilir (Vohra ve Tivari, 2020).

Şekillerin sınıflandırılabilmesi için kullanılan birçok yöntem vardır. Bu amaçla, şekil göstergeleri yaygın olarak kullanılır. Şekil göstergeleri, bir detayın (nesnenin) kendi geometrisinin ya da bu geometri üzerinden elde edilen yardımcı geometrilerin metrik büyüklükleri aracılığıyla detayların belirli şekilsel özelliklerinin sayısal olarak ifade edilmesini ve böylece birbirleriyle karşılaştırılabilirliklerini, benzerliklerinin ölçülebilmesini ve sınıflandırılabilirliklerini sağlayan ölçülerdir (Aktaş, 2012; Angel ve Ark., 2010; Basaraner ve Cetinkaya, 2017; Da Costa ve Cesar, 2009; Wentz, 1997).

Bir şekil göstergesinin sahip olması gereken özellikler şöyledir (Arkin ve Ark., 1991): (1) metrik olmalıdır, (2) öteleme, dönüklük ve ölçek değişiminden etkilenmemelidir, (3) hesaplaması kolay olmalıdır ve (4) sezgilerimizle uyusmalıdır. Şekil göstergeleri, uygun biçimde formüle edildiklerinde genellikle 0 ile 1 arasında değer üretirler. Çokgenlerin şekilleri ne kadar karmaşıkta şekil göstergelerinin ürettiği değerler 0'a o kadar yakın olur. Ters durumda, değerler 1 ya da 1'e yakın olur (Basaraner ve Cetinkaya, 2017).

Şekil göstergelerinin her biri şeklin belirli bir özelliğini ifade etmektedir. Dolayısıyla, çoğu zaman tek bir şekil göstergesi şekilleri sınıflandırmada yeterli olmaz (Aktaş, 2012; Angel ve Ark., 2010; Basaraner ve Cetinkaya, 2017; Wentz, 1997). Şekil analizine ilişkin çeşitli çalışmalar yapılmış olsa da bunların önemli bir kısmı raster verilere odaklanmıştır. Vektör veriler üzerine yapılan çalışmalar daha azdır (Ai ve Ark., 2013). Medda ve Ark. (1998) yerleşim alanı şekillerinin analizi için Boyce ve Clark'ın önerdiği şekil göstergesini esas alarak yeni bir gösterge geliştirmişlerdir. Zhao ve Stough (2005) parçalanmış veya boşluklu çeşitli çokgenler arasındaki şekil benzerliğini analiz etmek için örtüştürmeye dayalı bir uzanım göstergesi önermişlerdir. Ai ve Ark., (2013) bina detayları örneğinde Fourier dönüşümü temelli şekil tanıma ve eşleştirme yöntemi sunmuştur. Xu ve Ark. (2017) boşluk içeren çokgenlerin benzerliğini analiz etmek için açılar ve mesafeler gibi değişmezleri kullanarak karmaşık geometrik şekillerde boşluk dağılımını tanımlamak için konum çizgelerini ve ayrıca boşluklu çokgenler arasındaki benzerliği ölçmek için ise Fourier tanımlayıcıları ve konum çizgelerine dayalı bir yöntem önermişlerdir. Basaraner ve Cetinkaya (2017) binaların algısal şekil karmaşıklık düzeylerini (basit, orta ve karmaşık) CBS ortamında çok sayıda şekil göstergesi ve

dört farklı sınıflandırma şeması (eşit aralık, standart sapma, doğal kırılma ve kutu grafiklerine dayalı) çiftlerini kullanarak elde etmiş ve tek başlarına kullanıldığında kendileri tarafından önerilen iki yeni gösterge dahil olmak üzere az sayıda şekil göstergesinin iyi performans sergilediğini ve şekil karmaşıklık düzeyi arttıkça bu performansın azaldığını göstermişlerdir.

Bu bağlamda, şekil göstergelerinin birlikte kullanımının sınıflandırma performansını arttırabileceği öngörülebilir. Ayrıca, bu amaçla kullanılan sınıflandırma yöntemleri de önemlidir. Bu kapsamda, makine öğrenme algoritmaları bir potansiyele sahiptir. Literatürde vektör verilerle temsil edilen binaların şekil karmaşıklığının incelemesine ilişkin olarak şekil göstergeleri ve makine öğrenmesi sınıflandırma yöntemlerinin birlikte kullanıldığı bir çalışmaya rastlanmamıştır. Bu yaklaşımın sınıflandırma doğruluğunu artırabileceği öngörülebilir. Bu nedenle, bu çalışmada bina şekil karmaşıklık düzeylerinin (basit, orta ve karmaşık) otomatik olarak belirlenmesi için sık tercih edilen üç şekil göstergesi (dairesellik, dikdörtgensellik ve dışbükeylik) ve başarısı nedeniyle yaygın kullanılan iki topluluk öğrenme algoritmasının (rastgele orman ve gradyan artırma) sınıflandırma performanslarının incelenmesi amaçlanmaktadır.

2. Yöntem ve uygulama

Çalışma, üç aşamadan oluşmaktadır: (1) Eğitim (kıyas) verilerinin hazırlanması, (2) Şekil göstergelerin hesaplanması, (3) Topluluk öğrenmesi algoritmalarıyla detayların karmaşıklık düzeylerine göre sınıflandırılması ve sınıflandırma doğruluğunun değerlendirilmesi.

2.1. Kullanılan veriler

Bu çalışmada Almanya'nın Dresden şehrine ait OpenStreetMap (OSM) bina veri seti kullanılmıştır. Bu bölgenin seçilme nedeni, farklı karmaşıklık derecelerine sahip birçok binanın bir arada bulunmasıdır. Bu veri setinde yaklaşık 5000 adet bina bulunmaktadır. Bu veri seti içinden 300 adet bina deneysel çalışmada kullanılmıştır. Şekil göstergeleri hesaplanmadan önce, CBS araçları yardımıyla bina geometrilerindeki hatalar (gereksiz ya da çift noktalar ve kendini keme gibi) giderilmiştir. Bu veriler, sınıflandırma aşamasındaki çapraz doğrulama işlemi kapsamında eğitim ve test verisi olarak kullanılmıştır.

2.2. Eğitim (kıyas) verisinin hazırlanması

Bina detayları, genel itibariyle aşağıdaki hususlar dikkate alınarak görsel algı temelinde farklı sınıflara ayrıştırılabilir (Basaraner ve Cetinkaya, 2017):

- (1) kenar (nokta) sayısı,
- (2) kenar uzunluklarının homojenliği/heterojenliği,
- (3) iç açıların homojenliği/heterojenliği,
- (4) çokgenin büyüklüğüne göre sınırı boyunca girinti ve çıkıntıların büyüklüğü ve sayısı,
- (5) geometrideki simetriklik/asimetriklik.

Şekil algısına ilişkin bu hususlar dikkate alınarak, bina detayları için üç şekil karmaşıklık düzeyi (basit, orta, karmaşık) oluşturulmuş ve binalar buna göre etiketlenmiştir. Eğitim verisi, etiketlenmiş çok daha geniş bir veri seti içinden her sınıfa eşit sayıda (100'er) bina düşecek biçimde rastgele seçilmiştir (Şekil 1).



Şekil 1. Örnek bina detaylarının şekil karmaşıklık kategorileri

2.3. Şekil göstergelerinin hesaplanması

Şekil göstergeleri, detayların belirli şekilsel özelliklerini sayısal olarak ifade etmeye yarayan matematiksel ölçülerdir (Wentz, 1997). Bu amaçla, detayın kendi geometrisi ve/veya detayın geometrisi aracılığıyla elde edilen yardımcı geometrileri (dışbükey çokgen, minimum alanlı sınırlayıcı dikdörtgen, kenarlar, ağırlık merkezi vb.) üzerinden elde edilen metrik büyüklüklerden (alan, çevre, kenar uzunluğu, ağırlık merkezine olan mesafe, açı vb.) yararlanılır. Şekil analizi için kullanılan birçok şekil göstergesi mevcuttur. Yapılacak uygulamaya göre, şekillerin bazı özellikleri diğerlerinden daha önemli olabilir (Da Costa ve Cesar, 2009). Bu nedenle uygulamaya göre uygun şekil göstergeleri seçilmelidir. Bu çalışmada, çeşitli mekansal çalışmalarda (örn. Basaraner ve Cetinkaya, 2017; Fleischmann ve Ark., 2020; Oksanen, 2013; Yan ve Ark., 2019) şekil analizi için sıklıkla tercih edilen üç farklı şekil göstergesi (dairesellik, dışbükeylik ve dikdörtgensellik) kullanılmıştır.

2.3.1. Dairesellik

Dairesellik, bir çokgenin daireye benzerlik oranını belirler. Başka bir ifadeyle, çokgenin şekilsel kompaktlığını ölçer (Basaraner ve Cetinkaya, 2017; Burghardt ve Steiniger, 2005). Aynı düzlemde yer alan sabit bir noktaya eşit uzaklıktaki noktaların oluşturduğu geometri olan daire, en kompakt şekil olarak kabul edilir (Angel ve Ark., 2010). Bir çokgenin şekli, daireden ne kadar farklı ise, yani ne kadar düzensiz ve/veya ince uzun ise dairesellik değeri o kadar küçük olur. Bir çokgenin dairesellik değeri, o çokgenin alanının kendisine eşit çevre uzunluğuna sahip bir dairenin alanına oranlanmasıyla hesaplanır (Eşitlik 1). Daha açık bir ifadeyle, çokgenin karşılaştırılacağı dairenin alanının hesaplanması için yarıçapının bilinmesi gerekir. Bu amaçla, çokgenin çevre değeri doğrudan dairenin çevre değeri olarak alınır ve yarıçap, dairenin çevresini veren eşitlik kullanılarak elde edilir. Dolayısıyla, gerekli sadeleştirmeler yapıldığında dairesellik, çokgenin alanının ve çevresinin bir fonksiyonu olur. Bu eşitlikte, D dairesellik, A_{CG} ve A_{ECD} sırasıyla çokgen ve eşit çevreli dairenin alanlarını ve C_{CG} çokgenin çevresini ifade etmektedir. Daire şekilli çokgenlerin dairesellik değerleri, 1 olur. Çokgenlerin şekilleri, daireden ne kadar farklı ise dairesellik değerleri sıfıra o kadar yakın olur.

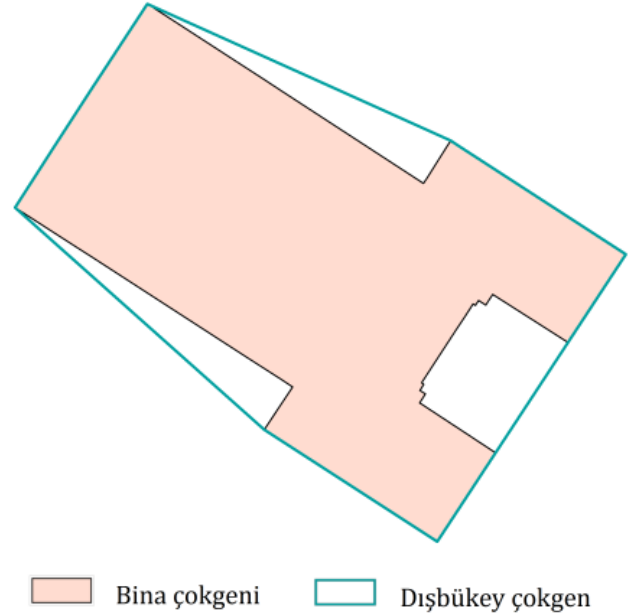
$$D = \frac{A_{CG}}{A_{ECD}} = \frac{4\pi A_{CG}}{C_{CG}^2} \quad (1)$$

2.3.2. Dışbükeylik

Dışbükeylik, bir çokgenin kendi minimum dışbükey çokgeninden (convex hull) olan farklılaşma oranını ortaya koyar (Şekil 2). Başka bir ifadeyle, bir çokgenin içe veya dışa doğru kıvrımlı olma derecesini gösterir (Basaraner ve Cetinkaya, 2017; Burghardt ve Steiniger, 2005; Ruas ve Holzapfel, 2003). Çokgenin alanının minimum dışbükey çokgeninin alanına oranıyla hesaplanır (Eşitlik 2). Bu eşitlikte, DB dışbükeylik şekil göstergesini, A_{CG} ve A_{MDC} sırasıyla çokgenin ve minimum

dışbükey çokgeninin alanlarını ifade etmektedir. Bu şekil göstergesi, bütün dışbükey çokgenler için 1 değerini üretir. Çokgenin şekli dışbükeylikten uzaklaştıkça, yani içbükeyliği arttıkça bu değer sıfıra yaklaşır.

$$DB = \frac{A_{CG}}{A_{MDC}} \quad (2)$$



Şekil 2. Bina çokgeni ve minimum dışbükey çokgeni

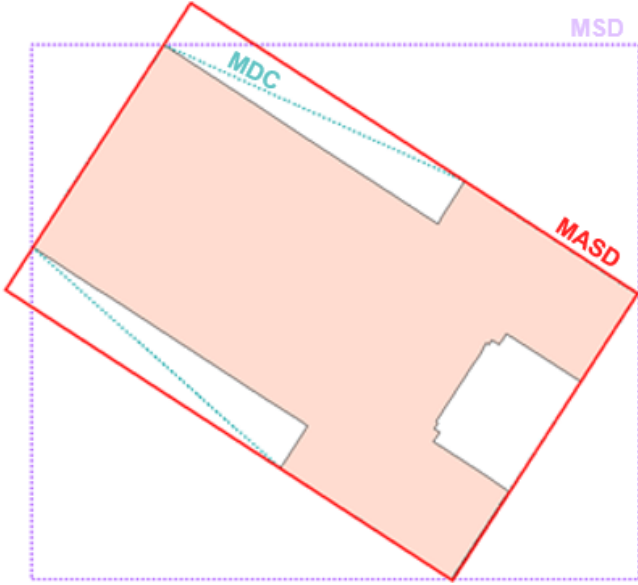
2.3.3. Dikdörtgensellik

Dikdörtgensellik, bir çokgenin şeklinin dikdörtgene (ve özel durumda kareye) benzerlik derecesini gösterir (Basaraner ve Cetinkaya, 2017; Rosin, 2000). Çokgenin alanının kendi minimum alanlı sınırlayıcı dikdörtgeninin (MASD) alanına oranıyla hesaplanır (Eşitlik 3). Bu eşitlikte, DD dikdörtgensellik, A_{CG} ve A_{MASD} sırasıyla çokgenin ve minimum alanlı sınırlayıcı dikdörtgeninin alanlarını ifade etmektedir. Dikdörtgen ve kare şekilli çokgenlerin dikdörtgensellik değerleri, 1 olur. Çokgenlerin şekilleri, dikdörtgen (veya özel durumda kareden) ne kadar farklı ise dikdörtgensellik değerleri sıfıra o kadar yakın olur.

$$DD = \frac{A_{CG}}{A_{MASD}} \quad (3)$$

MASD, bir çokgenin dışbükey çokgenine dayalı olarak hesaplanır (Toussaint, 1983). Dışbükey çokgen, sırasıyla her bir kenarı yatay eksene paralel olacak biçimde ilgili köşe noktasına dayalı olarak döndürülür ve her seferinde minimum sınırlayıcı dikdörtgeni (MSD) elde edilir, alan değeri ve ilgili köşe nokta numarası kaydedilir. Sonrasında en küçük alan değerini veren MSD, ilgili dışbükey çokgen kenarının yatayla yaptığı açı kadar ilgili köşe noktasına dayalı olarak ters yönde döndürülerek MASD elde edilir. MSD, çokgenin minimum ve maksimum koordinatlarından oluşturulan yatay dikdörtgendir. Bu nedenle çokgenin dönüklüğü MSD'yi ve dolayısıyla şekil gösterge değerini etkiler. MASD ile bu olumsuzluk giderilir. Daha açık bir ifadeyle, farklı dönüklüklere ve

aynı şekle sahip çokgenlerin MSD'leri farklı iken MASD'leri aynı olur. Şekil 3, dikdörtgensellik ile ilintili geometrileri göstermektedir. Çokgenlerin şekilleri MASD'lerinden ne kadar farklı ise dikdörtgensellik değerleri sıfıra o kadar yakın olur.



Şekil 3. Dikdörtgensellik ile ilintili geometriler: minimum dışbükey çokgen (MDC), minimum sınırlayıcı dikdörtgen (MSD) ve minimum alanlı sınırlayıcı dikdörtgen (MASD)

2.3.4. Bina Detayları ve Şekil Gösterge Değerleri

Bu bölümde bina detaylarına ilişkin olarak şekil karmaşıklık düzeylerine göre şekil göstergelerinin ürettiği değerler hakkında fikir verebilmek amacıyla örnek bina çokgenleri ve bu çokgenlerin şekil gösterge değerleri Tablo 1'de gösterilmiştir. Şekil göstergeleri, MapInfo Pro CBS yazılımı üzerinde MapBasic programlama diliyle geliştirilen bir eklenti yardımıyla hesaplanmıştır (Şekil 4).

Tablo 1. Örnek bina çokgenleri ve şekil gösterge değerleri

		Karmaşıklık düzeyi		
		Basit	Orta	Karmaşık
Şekil göstergeleri	D	0.785	0.418	0.187
	DB	1	0.829	0.612
	DD	1	0.775	0.487



Şekil 4. Bina detaylarının şekil gösterge değerleri ve hesaplamada kullanılan yazılım eklentisi

2.4. Topluluk öğrenmesi algoritmalarıyla detayların karmaşıklık düzeylerine göre sınıflandırılması

Bina detaylarının şekilsel karmaşıklık düzeylerine göre sınıflandırılması için genel olarak yüksek doğruluklu sonuçlar üretmesi (Breiman, 2001; Han ve Ark., 2012; Tattar, 2018) ve yaygın kullanımları (Witten ve Ark., 2016) nedeniyle topluluk öğrenmesi algoritmalarından rastgele orman ve gradyan artırma algoritmaları tercih edilmiştir. Sınıflandırma işlemi, veri bilimi ve makine öğrenmesi yazılımı olan RapidMiner ile gerçekleştirilmiştir.

Topluluk öğrenmesi, tek bir süper doğru modeli öğrenmeye çalışmak yerine, çok sayıda düşük doğruluklu modeli eğitmeye ve ardından yüksek doğrulukta bir üst model elde etmek için bu zayıf modeller tarafından üretilen tahminleri birleştirmeye odaklanan bir öğrenme paradigmasıdır (Burkov, 2019). Bu nedenle, topluluk sınıflandırıcıları genellikle yüksek doğrulukta modeller üretir (Burkov, 2020; Vukicevic ve Ark., 2013). Bütün topluluk öğrenme algoritmalarında olduğu gibi, eğitimde kullanılan ağaçların çeşitliliği ve sayısı ne kadar çok olursa, topluluğun hatası o kadar düşük olur (Breiman, 2001; Kotu ve Deshpande, 2019).

Akar ve Görmüş (2019) Gram Schmidt ve ana bileşen yöntemleri ile Göktürk 2 multispektral, Göktürk 2 pankromatik ve Hyperion EO-1 uydu görüntülerini kaynaştırarak daha yüksek doğrulukta arazi kullanım haritasının üretilmesini amaçlamışlardır. Sınıflandırma için rastgele orman ve destek vektör makineleri algoritmalarını kullanmışlar ve en yüksek doğruluğun rastgele orman algoritması ile elde edildiğini gözlemlemişlerdir. Görmüş ve Akar (2019) hiperspektral hava fotoğraflarının boyutunu indirgeme yöntemlerinin performanslarını değerlendirmek için rastgele orman ve destek vektör makineleri algoritmalarını kullanmışlardır. Bu bağlamda, rastgele orman algoritması ile biraz daha yüksek doğruluklu sonuçlar elde etmişlerdir. Üstüner ve Ark., (2020) yeni nesil topluluk öğrenme algoritmaları olan aşırı gradyan artırma ve hafif gradyan artırma makineleri algoritmalarının tarım ürünlerinin sınıflandırılmasındaki performanslarını karşılaştırmışlardır. Hafif gradyan artırma algoritmasıyla daha yüksek doğruluklu sonuçlar elde etmişlerdir. Ayrıca, bu bağlamda parametre seçiminin gücüne de değinmişlerdir. Zeybek (2020) nokta bulutundan ağaç gövdesinin otomatik çıkarımında lojistik regresyon, doğrusal ayrıştırma analizi, rastgele orman ve destek vektör makinesi algoritmalarının performanslarını analiz etmiştir. Bu bağlamda, sınıflandırma doğruluğu en yüksek yöntemin rastgele orman algoritması olduğunu saptamışlardır. Zeybek (2021) İHA görüntüleri ile elde edilen görüntülerden üretilen nokta bulutunun, radyometrik ve geometrik (kovaryans) özelliklerine göre sınıflandırılması için rastgele orman algoritmasıyla %96 genel doğruluk elde etmiştir.

Rastgele orman ve gradyan artırma algoritmalarında girdi olarak kullanılan özellik (öznitelik) değerlerinin normalleştirilmesinin veya standartlaştırılmasının sonuçlara anlamlı bir etkisi yoktur (Tattar, 2018).

Öğrenme aşamasında veri setini eğitim ve test verileri biçiminde manuel olarak bölmek yerine çapraz

doğrulama yöntemi kullanılmıştır. Bu yöntem, veri setini rastgele olarak eşit parçalara bölerek her seferinde bir parçayı test için ayırıp diğer parçalardan eğitim işlemi yapar ve son olarak her modelden elde edilen başarı oranlarının ortalamalarını alır (Bonaccorso, 2020). Böylelikle, aşırı öğrenme (over-fitting) sorunu önlenmiş ve bütün veriler hem eğitim hem de test için kullanılmış olur (Euler, 2014). Bu çalışmada, Han ve Ark. (2012) tarafından önerilen k-katlamalı ($k = 10$) çapraz doğrulama tekniği kullanılmıştır.

2.4.1. Rastgele orman algoritması

Rastgele orman algoritması, torbalama (bagging) tekniğine dayanan karar ağacı tabanlı bir topluluk öğrenme algoritmasıdır (Bonaccorso, 2020; Breiman, 2001).

Bu algoritmada, topluluktaki sınıflandırıcıların her birinin bir karar ağacı sınıflandırıcısı olduğu ve böylece sınıflandırıcıların oluşturduğu grubun bir "orman" olduğu kabul edilir. Ayrı karar ağaçları, bölünmeyi belirlemek için her düğümde rastgele bir özellik (öznitelik) seçimi kullanılarak oluşturulur. Başka bir ifadeyle, her ağaç, ormandaki tüm ağaçlar için bağımsız olarak ve aynı dağılımla örneklenen rastgele bir vektörün değerlerine bağlıdır. Sınıflandırma sırasında, her karar ağacı, o ağacın eğitiminde kullanılmayan örneklerin sınıflandırılmasına oy verir ve en çok oy alan sınıf döndürülür (Han ve Ark., 2012; Kotu ve Deshpande, 2019).

Rastgele orman algoritmasının temel adımları şöyledir (Yang, 2019):

- 1: Verileri yükle ve kullanıma hazırla.
- 2: Döngüyü başlat (her bir ağaç için).
- 3: D adet özellikten rastgele m adet özellik seç.
- 4: T adet farklı ağaç oluşturmak için verileri örnekle.
- 5: Karar düğümlerini iteratif olarak bölerek her bir ağacı oluştur.
- 6: Çoklu ağaçlara dayalı olarak herhangi bir test özelliğini tahmin et.
- 7: Çoğunluk oylamasını kullanarak sınıflandır.
- 8: Döngüyü bitir.

Rastgele orman, en yaygın olarak kullanılan ve en etkili topluluk öğrenme algoritmalarından biridir. Etkili olmasının nedeni, orijinal veri kümesinin birden çok örneğini kullanarak son modelin varyansının azaltılmasıdır. Düşük varyans, düşük aşırı öğrenme anlamına gelmektedir. Model, veri kümesindeki küçük değişimleri açıklamaya çalıştığında aşırı öğrenme gerçekleşir, çünkü veri kümesi, modellenmeye çalışılan olgunun tüm olası örneklerine ilişkin yığın yalnızca küçük bir örneklemdir. Eğitim setinin değiştirilmesiyle birden fazla rastgele örnek oluşturarak istenmeyen unsurların (gürültü, aykırı değerler ve fazla veya az temsil edilen örnekler) etkisi azaltılır (Burkov, 2019).

Uygulama aşamasında, algoritma çalıştırılmadan önce çeşitli hiperparametrelerin (ağaç sayısı, maksimum derinlik vb.) girilmesi gerekmektedir. Yazılım parametre optimizasyonunu mümkün kılmaktadır. Böylece uygun parametre aralıkları seçilerek, bu aralıklarda en iyi sonuç

üreten parametre değerleri iteratif olarak belirlenebilmektedir. Bu bağlamda, bazı hiperparametreler için uygun minimum ve maksimum değerler, literatür araştırması ve denemeler sonucunda kararlaştırılmıştır.

Önemli hiperparametrelerden biri, ağaç sayısıdır. Ağaç sayısı arttıkça ve çeşitlendikçe topluluğun hatası bir dereceye kadar azalır. Bu sınıra ulaştıktan sonra ağaç sayısındaki artışın performansa etkisi önemsiz sayılabilecek kadar az olabilir ve hesaplama süresinin gereksiz uzamasına neden olur (Breiman, 2001). Bu nedenle, ağaç sayısı ve diğer hiperparametrelerin optimizasyonu, ızgaralı arama (grid search) yöntemiyle farklı kombinasyonlar denenerek sağlanır.

Ağaç sayısının küçük veri setleri için 50, büyük veri setleri için ise 500 ve daha fazla seçilmesinin uygun olacağı belirtilmiştir (Zeybek, 2021). Oshiro ve ark., (2012) ise 64-128 aralığında olmasını önermektedir. Bu aralık biraz daha genişletilerek denemeler yapılmıştır. Bununla birlikte, optimum değer, önerilen aralıkta yer almıştır. Maksimum derinlik parametresi, 5-20 aralığında seçilmiştir. Yani, varsayılan değerler yerine daha geniş bir değer aralığı seçilmiştir. Yazılım, bu değer aralıkları içinde optimizasyon yaparak hiperparametrelerin en uygun değerlerini belirlemektedir. Burada, belirli bir aralık değerinin yani minimum ve maksimum değerlerin girilmesinin nedeni, işlem süresinin kısaltılmasıdır. Doğrulukların istenen düzeyde olmaması durumunda seçilen değer aralıkları gözden geçirilebilir.

2.4.2. Gradyan artırma algoritması

Gradyan artırma, artırma (boosting) tekniğine dayanan karar ağacı tabanlı bir topluluk öğrenme algoritmasıdır (Freeman, 2001). Gradyan artırma algoritması, tahminleri rastgele orman algoritmasına göre farklı bir teknik kullanarak yapar. Burada, ağaç oluşturma sürecini paralelleştirmek yerine, tahminleri elde etmek için sıralı bir yaklaşım benimser. Gradyan, bir model oluşturduktan sonra elde edilen hatayı veya artığı ifade eder. Gradyan artırmada, her karar ağacı bir önceki karar ağacının hatasını tahmin eder ve böylece hatayı (gradyan) kademeli olarak azaltır (Ayyadevara, 2018).

Gradyan artırma algoritmasının temel adımları şöyledir (Ayyadevara, 2018):

- 1: Basit bir karar ağacıyla tahminleri başlat.
- 2: Artıkları (gerçek değer ve tahmin değeri farkını) hesapla.
- 3: Tüm bağımsız değerlere dayalı olarak artıkları tahmin eden başka bir sığ karar ağacı oluştur.
- 4: Orijinal tahmini, öğrenme oranıyla çarpılan yeni tahminle güncelle.
- 5: Belirli bir iterasyon sayısı için 2'den 4'e kadar olan adımları tekrarla (iterasyon sayısı, ağaç sayısı olacaktır).

Artırma algoritmalarının ardındaki genel düşünce, zayıf öğrenenleri güçlü öğrenenlere dönüştürmektir. Zayıf bir öğrenici, rastgele tahminde bulunmaktan pek de iyi değildir. Ancak zayıf başlangıcın arkasında bir amaç vardır. Bu genel düşünceye dayalı olarak, güçlü bir temel

model kurarak değil, yinelemeli hata düzeltmeye odaklanarak işler hızlandırılır. Temel model çok güçlüyse, öğrenme süreci zorunlu olarak sınırlıdır ve bu nedenle, artırma modellerinin arkasındaki genel stratejiye zarar verir. Zayıf öğrenenler, yüzlerce iterasyon yoluyla güçlü öğrenicilere dönüştürülür. Gradyan artırma, her yeni ağaç için hatalara bakar ve ardından tamamen bu hataların etrafında yeni bir ağaç oluşturur. Yeni ağaç, zaten doğru olan tahminleri dikkate almaz. Gradyan artırma, başarılı sonuçlar üretme açısından en iyi makine öğrenimi algoritmalarından biri olmuştur (Wade, 2020).

Bu algoritma için ise ağaç sayısının 40-70, maksimum derinlik parametresinin ise 5-8 aralığında seçilmesi önerilmektedir (URL-1). Uygulamada bu öneriler de dikkate alınarak çeşitli denemeler yapılmış ve bu değerler uygun bulunmuştur. Yazılım, yine bu değer aralıkları içinde optimizasyon ile en iyi sonucu veren değeri elde etmektedir.

2.4.3. Doğruluk değerlendirme

Sınıflandırma doğruluğunun değerlendirilmesi, hata (karışıklık) matrisine dayalı olarak gerçekleştirilir (Tablo 2). Hata matrisi, sınıflandırıcının farklı kategorileri ne kadar iyi tahmin ettiğini analiz etmek için kullanışlı bir araçtır (Han ve Ark., 2012). Bu matris, belirli bir kategoriye atanan birim sayısını "gerçeğe" (referans verilere) kıyasla ifade eden karesel sayı dizisinden oluşur. Bu yöntem, örneğin sınıflandırılmış bir uydu görüntüsünün veya tematik haritaların doğruluğunu değerlendirmek için uygulanabilir (Shi, 2010).

Genel doğruluk, en yaygın olarak rapor edilen doğruluk değerlendirme istatistiğidir. Ancak, sadece genel doğruluğu vermek yeterli değildir. Diğer doğruluk ölçütlerinin gerektiği gibi hesaplanabilmesi ve sınıflar arasındaki herhangi bir karışıklığın açıkça sunulabilmesi ve anlaşılabilmesi için matrisin tamamının sunulması önemlidir. Üretici ve kullanıcı doğrulukları, tüm matris için genel sınıflandırma doğruluğunun hesaplanmasına ek olarak bireysel sınıf doğruluklarını belirlemek için hesaplanabilir (Congalton ve Green, 2019). F_1 -skor ise üretici doğruluğu ve kullanıcı doğruluğu parametrelerinin harmonik ortalaması alınarak hesaplanır (Han ve Ark., 2012). Kappa parametresi, iki veri setinin gözlemlenen uyumu ile tesadüfen olan uyumu arasında farkın ölçümünü sağlar (Shi, 2010). Tüm hata matrisi için kappa parametresini hesaplamaya ek olarak, matris içindeki her bir kategori için uyuma bakmak faydalı olabilir. Bireysel kategori uyumu, koşullu kappa parametresi kullanılarak test edilebilir (Congalton ve Green, 2019). Bu değerler ne kadar yüksek ise sınıflandırma doğruluğu o oranda başarılıdır.

Bu kapsamda, genel doğruluk (Gnl_Dg), kappa (κ), üretici doğruluğu (hassasiyet) (Urt_Dg), kullanıcı doğruluğu (kesinlik) (Kul_Dg), F_1 -skor ve koşullu kappa (Kos_k) parametreleri sırasıyla 4-9 eşitlikleriyle hesaplanmıştır (Congalton ve Green, 2019; Han ve Ark., 2012).

Tablo 2. Hata matrisi (Congalton ve Green, 2019)

		j = Sütunlar (Gerçek / Referans) toplamı			Satır
		1	2	k	n _{i+}
i = Satırlar (Tahmin / Sınıflandırma)	1	n ₁₁	n ₁₂	n _{1k}	n ₁₊
	2	n ₂₁	n ₂₂	n _{2k}	n ₂₊
	k	n _{k1}	n _{k2}	n _{kk}	n _{k+}
Sütun toplamı		n ₊₁	n ₊₂	n _{+k}	n
n _j					

$$Gnl_Dg = \frac{\sum_{i=1}^k n_{ii}}{n} \times 100 \quad (4)$$

$$\kappa = \frac{n \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i+} n_{+i}}{n^2 - \sum_{i=1}^k n_{i+} n_{+i}} \quad (5)$$

$$Urt_Dg = \frac{n_{jj}}{n_{+j}} \times 100 \quad (6)$$

$$Kul_Dg = \frac{n_{ii}}{n_{i+}} \times 100 \quad (7)$$

$$F_1 - skor = \frac{2 \times Urt_Dg \times Kul_Dg}{Urt_Dg + Kul_Dg} \quad (8)$$

$$Kos_k = \frac{nn_{ii} - n_{i+} n_{+i}}{nn_{i+} - n_{i+} n_{+i}} \quad (9)$$

3. Bulgular ve tartışma

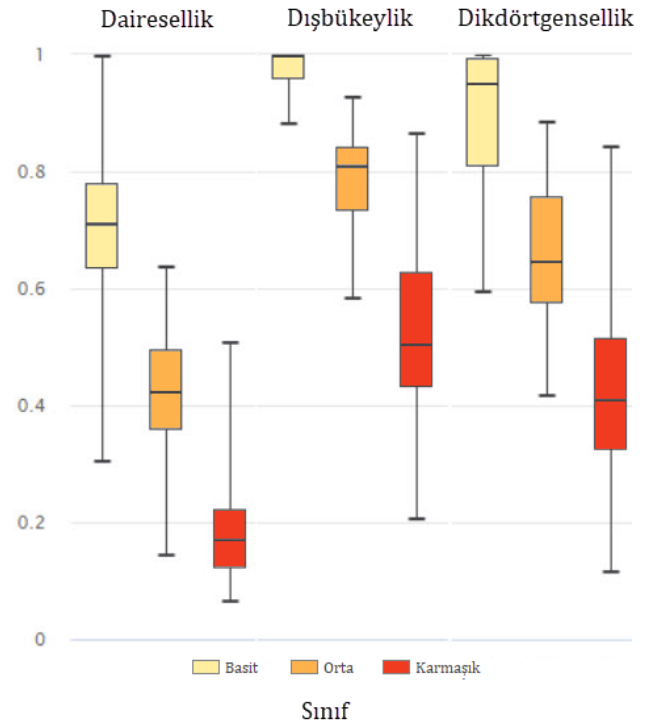
Çalışmada kullanılan bina veri seti için hesaplanan şekil göstergelerine ilişkin betimleyici istatistikler, Tablo 3'te verilmiştir.

Tablo 3. Şekil göstergelerinin istatistiksel değerleri

	D	DB	DD
Min.	0.064	0.204	0.115
Maks.	0.997	1	1
Ort.	0.431	0.762	0.659
St. Sapma	0.230	0.212	0.232
Med.	0.414	0.810	0.658

Ayrıca eğitim verisindeki görsel algıya dayalı olarak belirlenen sınıflara göre şekil göstergelerinin değer dağılımı, Şekil 5'te kutu grafiğiyle gösterilmektedir. Buna göre, sınıf (şekil karmaşıklık kategorileri) temelinde şekil gösterge değerlerinin %50'lik kısmının gayet net olarak ayrıştırılabildiği anlaşılmaktadır. Bununla birlikte, değerlerin tamamı dikkate alındığında tüm kategoriler arasında çakışmaların var olduğu görülmektedir.

Rastgele orman ve gradyan artırma algoritmalarıyla gerçekleştirilen sınıflandırma işlemleri sonucunda elde edilen hata matrisleri sırasıyla Tablo 4 ve 5'te, bu matrise dayalı hesaplanan doğruluk parametrelerinin değerleri ise sırasıyla Tablo 6 ve 7'de verilmiştir.



Şekil 5. Şekil göstergelerinin şekil karmaşıklık kategorilerine göre değer dağılımlarını gösteren kutu grafiği

Çoklu karar ağaçları üreten rastgele orman ve gradyan artırma algoritmalarından karar ağacı örnekleri sırasıyla Şekil 6 ve Şekil 7'de verilmiştir. Gradyan artırma algoritmasından elde edilen karar ağaçları sınıf temelli olup yorumlanması güçtür. Rastgele orman algoritmasından elde edilen karar ağacının yorumlanmasının daha kolay olduğu görülmektedir.

Tablo 4. Rastgele orman algoritması ile sınıflandırma sonucu elde edilen hata matrisi

		Gerçek			Satır toplamı
		Basit	Orta	Karmaşık	
Tahmin	Basit	98	2	0	100
	Orta	2	93	11	106
	Karmaşık	0	5	89	94
	Sütun toplamı	100	100	100	300

Tablo 5. Gradyan artırma algoritması ile sınıflandırma sonucu elde edilen hata matrisi

		Gerçek			Satır toplamı
		Basit	Orta	Karmaşık	
Tahmin	Basit	98	2	0	100
	Orta	1	87	8	96
	Karmaşık	1	11	92	104
	Sütun toplamı	100	100	100	300

Tablo 6. Rastgele orman algoritması ile sınıflandırma sonucu elde edilen doğruluk parametrelerinin değerleri

Rastgele Orman				
Gen_Dg		% 93.33		
κ		0.900		
Sınıf	Urt_Dg	Kul_Dg	F ₁ -skor	Kos_κ
Basit	0.98	0.98	0.98	0.97
Orta	0.93	0.88	0.90	0.82
Karmaşık	0.89	0.95	0.92	0.92

Tablo 7. Gradyan artırma algoritması ile sınıflandırma sonucu elde edilen doğruluk parametrelerinin değerleri

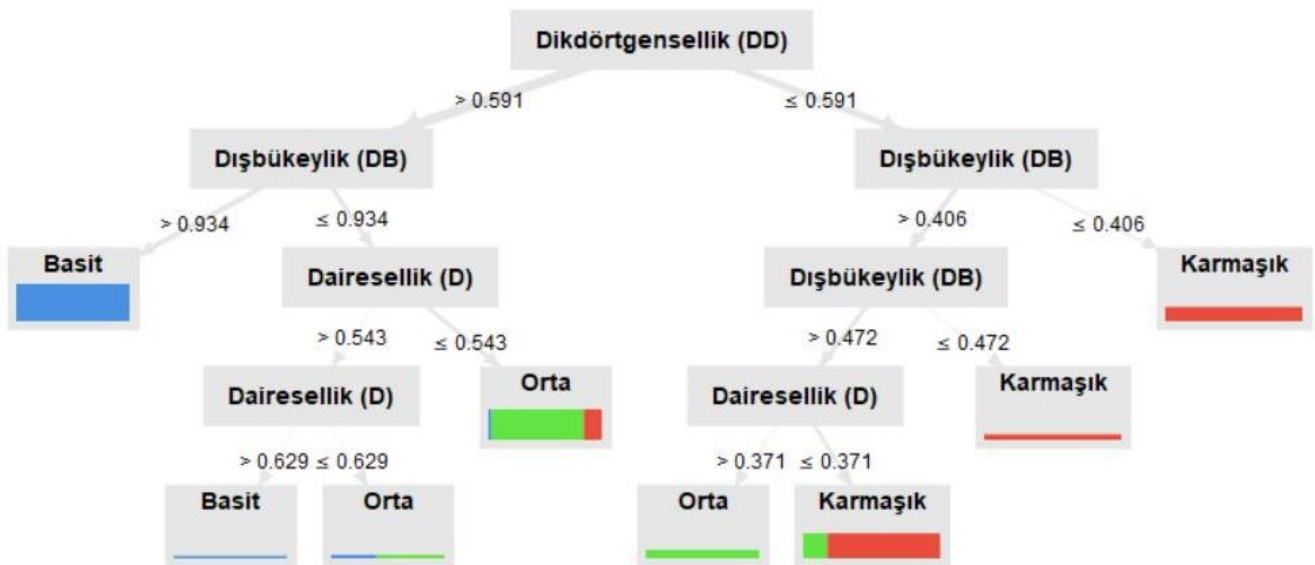
Gradyan Artırma				
Gen_Dg		%92.33		
κ		0.885		
Sınıf	Urt_Dg	Kul_Dg	F ₁ -skor	Kos_κ
Basit	0.98	0.98	0.98	0.97
Orta	0.87	0.91	0.89	0.86
Karmaşık	0.92	0.88	0.90	0.83

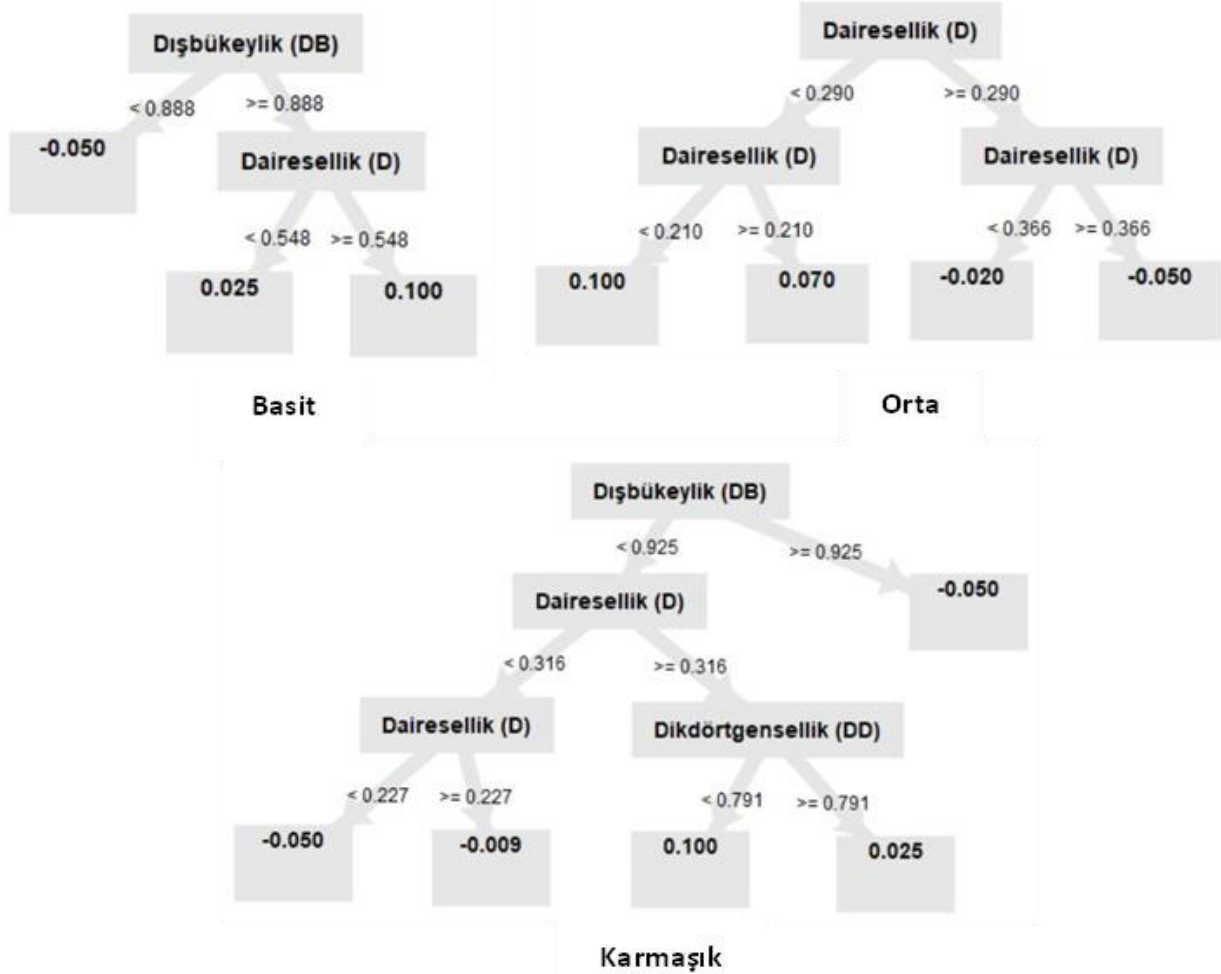
Rastgele orman algoritması, gradyan artırma algoritmasına göre çok az daha iyi bir genel performans göstermiştir. Her iki algoritma ile “basit” şekil karmaşıklık düzeyi, “orta” ve “karmaşık” şekil karmaşıklık düzeylerine göre daha başarılı biçimde sınıflandırılmıştır.

Çalışmanın sınırlılığı olarak üç husus belirtilebilir. İlki eğitim (kıyas) verisinin oluşturulmasıdır. Bu işlem görsel algıya dayalı olduğu için kısmen öznellik içermektedir.

Fakat bu bağlamda önceki literatürde yer alan genel ilkelere mümkün olduğunca uyularak bu potansiyel olumsuzluğun en aza indirgenmesine çalışılmıştır. İkinci olarak şekil göstergelerinin şekilleri tanımlama beceresinin kısıtlı olmasıdır. Başka bir ifadeyle, her şekil göstergesi şeklin belirli bir özelliğine odaklanmaktadır. Dolayısıyla, çoğu durumda tek başlarına şekilleri birbirinden ayırt etmede yeterince etkili olamamaktadırlar. Bu kapsamda, özellikle sınıf sayısının artması halinde daha fazla şekil göstergesinin katkı sağlayabileceği öngörülebilir. Son olarak, makine öğrenmesi algoritmalarının çok sayıda parametreye dayalı olması ve bu parametrelerin seçilmesindeki güçlükten bahsedilebilir. Bu problemi en aza indirmek amacıyla bazı temel hiperparametreler için önerilen değer aralıkları esas alınmış ve çeşitli ön denemeler yapılarak uygun aralıklara karar verilmiştir. Bu kapsamda, elde edilen doğruluk değerleri yönlendirici olmuştur. Sonrasında, yazılım bu aralıklar içinde en yüksek doğruluk veren parametreleri optimizasyonla belirlemektedir.

Basaraner ve Cetinkaya (2017) dairesellik, dışbükeylik ve dikdörtgensellik şekil göstergeleri ve kutu grafikleri yardımıyla belirlenen eşik değerlere dayalı sınıflandırma yaklaşımıyla sırasıyla %73.8 ($\kappa = 0.491$), %88.3 ($\kappa = 0.784$), %85.5 ($\kappa = 0.736$) genel doğruluk elde etmişlerdir. Bu şekil göstergeleri ile kullandıkları diğer sınıflandırma şemaları daha düşük doğruluk üretmiştir. Bu çalışma da ise, tekli şekil göstergesi kullanımına dayalı bulgulara göre şekil göstergelerinin birlikte hem rastgele orman ve hem de gradyan artırma topluluk öğrenme algoritmalarıyla kullanılmasının sınıflandırma doğruluğunu artırdığı gösterilmiştir.

**Şekil 6.** Rastgele orman algoritmasıyla elde edilen karar ağaçlarından bir örnek



Şekil 7. Gradyan artırma algoritmasıyla elde edilen karar ağaçlarından bir örnek

4. Sonuçlar

Bu makale, şekil göstergeleri ve topluluk öğrenmesi algoritmalarıyla bina detaylarının şekil karmaşıklık düzeylerinin otomatik olarak belirlenmesini ele almıştır. Bu kapsamda, önceki çalışmalarda başarılı bulunan dairesellik, dışbükeylik ve dikdörtgensellik şekil göstergeleri ile birlikte rastgele orman ve gradyan artırma topluluk öğrenmesi algoritmaları kullanılmıştır. Bu yaklaşım, OSM'den elde edilen 300 adet bina üzerinde test edilmiştir. Sınıflandırma doğruluğu incelemesi için eğitim verisi oluşturmak üzere tüm binalar şekil karmaşıklık düzeylerine göre basit, orta ve karmaşık olarak etiketlenmiştir. Deneysel çalışma sonucu, her iki algoritmayla oldukça yüksek doğruluk düzeyleri elde edilmiştir. Rastgele Orman algoritmasının çok az daha iyi sonuç verdiği gözlemlenmiştir. Dolayısıyla, birden fazla şekil göstergesinin, uygun topluluk öğrenmesi algoritmalarıyla kullanımının bina şekil karmaşıklık düzeylerini otomatik olarak belirlemede oldukça etkin olduğu ve önceki çalışmalarda yer alan tekli şekil göstergesi ve CBS sınıflandırma şemaları kullanımına göre performansı artırdığı görülmüştür.

Her iki algoritmadan elde edilen çoklu karar ağaçlarına dayalı modellerin farklı bina veri setlerinin sınıflandırılmasında kullanılabileceği öngörülebilir.

Gelecek çalışmalarda, daha fazla ve farklı şekilsel özellikleri ölçen şekil göstergeleri kullanılarak değerlendirmeler yapılabilir. Bu kapsamda, farklı makine

öğrenmesi algoritmaları ve istifleme (stacking) yöntemi de test edilebilir. Ayrıca, şekil karmaşıklık kategorilerinin sayısı artırılabilir ve uygulama daha kapsamlı bir veri setiyle gerçekleştirilebilir.

Araştırmacıların katkı oranı

Hüseyin Safa Duman: Literatür taraması, Yöntem, Uygulama, Makale yazımı; **Melih Başaraner:** Literatür taraması, Tasarım, Yöntem, Makale yazımı

Çatışma beyanı

Herhangi bir çıkar çatışması bulunmamaktadır.

Kaynakça

- Ai T, Cheng X, Liu P & Yang M (2013). A shape analysis and template matching of building features by the fourier transform method. *Computers, Environment and Urban Systems*, 41, 219–233.
- Akar Ö & Görmüş E T (2019). Göktürk-2 ve Hyperion EO-1 Uydu Görüntülerinden Rastgele Orman Sınıflandırıcısı ve Destek Vektör Makineleri ile Arazi Kullanım Haritalarının Üretilmesi. *Geomatik*, 4(1), 68-81.
- Aktaş M (2012). Shape Descriptors. Doktora Tezi, Exeter Üniversitesi, Bilgisayar Bilimleri Bölümü, İngiltere.

- Angel S, Parent J & Civco L (2010). Ten compactness properties of circles: measuring shape in geography. *The Canadian Geographer / Le Géographe canadien*, 54(4), 441-461.
- Arkin E, Chew L P, Huttenlocher D P, Kedem K & Mitchell J S B (1991). An efficiently computable metric for comparing polygonal shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3), 209-216.
- Basaraner M & Cetinkaya S (2017). Performance of shape indices and classification schemes for characterising perceptual shape complexity of building footprints in GIS. *International Journal of Geographical Information Science*, 31(10), 1952-1977.
- Basaraner M (2020). Geometric and semantic quality assessments of building features in OpenStreetMap for some areas of Istanbul. *Polish Cartographical Review*, 52(3), 94-107.
- Başaraner M (2005). Nesne Yönelimli Coğrafi Bilgi Sistemi Ortamında Orta Ölçekli Topografik Haritalar İçin Bina ve Yerleşim Alanlarının Otomatik Genelleştirilmesi. Doktora Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Bonaccorso G (2020). *Mastering Machine Learning Algorithms*, 2nd edn. Packt Publishing, Birmingham, 798 s.
- Breiman L (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Burghardt D & Steiniger S (2005). Usage of principal component analysis in the process of automated generalisation. *Proceedings of 22nd International Cartographic Conference*, La Coruna, Spain.
- Burkov A (2019). *The Hundred-Page Machine Learning Book*. Andriy B., 159 s.
- Burkov A (2020). *Machine Learning Engineering*. True Positive Inc., Quebec City, 310 s.
- Congalton R G & Green K (2019). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 3rd edn. CRC Press, Boca Raton, FL, 328 s.
- Çetinkaya S (2014). Kartografik Genelleştirmede Bina Dizilimlerinin Karakterizasyonu ve Yorumlanmasına İlişkin Yeni Yaklaşımlar. Doktora Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Da Costa F & Cesar R (2009). *Shape Classification and Analysis: Theory and Practice*, 2nd edn. CRC Press, Boca Raton, FL, 685 s.
- Demetriou D, See L & Stillwell J (2013). A parcel shape index for use in land consolidation planning. *Transactions in GIS* 17(6), 861-882.
- Euler T (2014). Who wants my product? Affinity-based marketing. M. Hofmann ve R. Klinkenberg (ed.) *Data Mining Use Cases and Business Analytics Applications*. Chapman & Hall/CRC Press, Boca Raton, FL, 77-96.
- Fleischmann M, Romice O & Porta S (2020). Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities. *Environment and Planning B: Urban Analytics and City Science*. DOI: 10.1177/2399808320910444
- Frenkel A & Ashkenazi M (2008). Measuring urban sprawl: how can we deal with it? *Environment and Planning B: Planning and Design*, 35(1), 56-79.
- Friedman J H (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29, 1189-1232.
- Görmüş E T & Akar Ö (2019). Headwall Hyperspec VNIR kamerası ile elde edilen hiperspektral hava fotoğrafı için boyut indirgeme yöntemlerinin performanslarının analizi. *Geomatik*, 4(3), 201-214.
- Grabler F, Agrawala M, Sumner W & Pauly M (2008). Automatic generation of tourist maps. *ACM Transactions on Graphics* 27(3), 1-11.
- Han J, Micheline K & Jian P (2012). *Data Mining: Concepts and Techniques*. 3rd edn. Morgan Kaufmann, Cambridge, MA, 744 s.
- Huang H, Kieler B & Sester M (2013). Urban building usage labeling by geometric and context analyses of the footprint data. *Proceedings of 26th International Cartographic Conference*, 25-30 August, Dresden, Germany.
- Kotu V & Deshpande B (2019). *Data Science: Concepts and Practice*, 2nd ed. Morgan Kaufmann Publishers, Cambridge, MA, 568 s.
- Li W, Goodchild F & Church R (2013). An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. *International Journal of Geographical Information Science*, 27(6), 1227-1250.
- Medda F, Nijkamp P & Rietveld P (1998). Recognition and classification of urban shapes. *Geographical Analysis*, 30(3), 304-314.
- Oksanen T (2013). Shape-describing indices for agricultural field plots and their relationship to operational efficiency. *Computers and Electronics in Agriculture*, 98, 252-259.
- Oshiro T M, Perez P S & Baranauskas J A (2012). How many trees in a Random Forest? P. Perner (ed.) *Machine Learning and Data Mining in Pattern Recognition- MLDM 2012*. *Lecture Notes in Computer Science*, vol. 7376. Springer, Berlin, Heidelberg, 154-168.
- Rosin P L (2000). Measuring shape: ellipticity, rectangularity, and triangularity. *Proceedings 15th International Conference on Pattern Recognition (ICPR-2000)*, Vol. 1, 952-955.
- Ruas A & Holzapfel F (2003). Automatic characterisation of building alignments by means of expert knowledge. *Proceedings of 21st International Cartographic Conference*, Durban, South Africa, 1604-1615.
- Shi W (2010). *Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses*. CRC Press, Boca Raton, FL, 454 s.
- Tattar P N (2018). *Hands-On Ensemble Learning with R*. Packt Publishing, Birmingham, 376 s.
- Toussaint G (1983). Solving geometric problems with the "Rotating Calipers". *Proceedings of IEEE MELECON'83*, A10.02/1-4.
- URL-1:
<https://www.analyticsvidhya.com/blog/2016/02/complete-guide-parameter-tuning-gradient-boosting-gbm-python/> Erişim Tarihi: 05.04.2021.
- Üstüner M, Abdikan S, Bilgin G & Şanlı F B (2020). Hafif gradyan artırma makineleri ile tarımsal ürünlerin sınıflandırılması. *Türk Uzaktan Algılama ve CBS Dergisi*, 1(2), 97-105.

- Vohra R & Tiwari K C (2020). Spatial shape feature descriptors in classification of engineered objects using high spatial resolution remote sensing data. *Evolving Systems*, 11(4), 647–660.
- Vukicevic M, Jovanovic Z, Delibašić B & Suknovic M (2013). Recommender system for selection of the right study program for higher education students. M. Hofmann ve R. Klinkenberg (ed.) *Data Mining Use Cases and Business Analytics Applications*. Chapman & Hall/CRC Press, Boca Raton, FL, 145-156.
- Wade C (2020). *Hands-On Gradient Boosting with XGBoost and scikit-learn: Perform accessible machine learning and extreme gradient boosting with Python*. Packt Publishing, Birmingham, 310 s.
- Wentz A (1997). Shape analysis in GIS. *Proceedings of Auto-Carto 13*, 204–213.
- Witten I H (2017). *Data Mining: Practical Machine Learning Tools and Techniques*. 4nd edn. Morgan Kaufmann, San Francisco, CA, 654 s.
- Xu Y, Xie Z, Chen Z & Wu L (2017). Shape similarity measurement model for holed polygons based on position graphs and Fourier descriptors. *International Journal of Geographical Information Science*, 31(2), 253-279.
- Yan X, Ai T, Yang M & Yin H (2019). A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150, 259-273.
- Yang X (2019). *Introduction to Algorithms for Data Mining and Machine Learning*. Academic Press, Cambridge, MA, 188 s.
- Zeybek M (2020). El-tipi LiDAR nokta bulutundan tek ağaç gövdesinin otomatik çıkarımında istatistiksel sınıflandırma algoritmalarının performans analizi. *Artvin Çoruh Üniversitesi Orman Fakültesi Dergisi*, 21 (2), 200-213.
- Zeybek M (2021). Classification of UAV point clouds by random forest machine learning algorithm. *Turkish Journal of Engineering*, 5(2),48-57.
- Zhao Z & Stough R R (2005). Measuring similarity among various shapes based on geometrical matching. *Geographical Analysis*, 37(4), 410–422.
- Zhong Y, Lin A, He L, Zhou Z & Yuan M (2020). Spatiotemporal dynamics and driving forces of urban land-use expansion: a case study of the Yangtze River Economic Belt, China. *Remote Sensing*, 12(2), 287.
- Zhou X, Chen Z, Zhang X & Tinghua A (2018). Change detection for building footprints with different levels of detail using combined shape and pattern analysis. *ISPRS International Journal of Geo-Information*, 7(10), 406–430.



© Author(s) 2022. This work is distributed under <https://creativecommons.org/licenses/by-sa/4.0/>