

Hafif Gradyan Artırma Makineleri ile Tarımsal Ürünlerin Sınıflandırılması

Crop Classification Using Light Gradient Boosting Machines

Mustafa Üstüner^{1*}, Saygın Abdikan², Gökhan Bilgin³, Füsün Balık Şanlı⁴

¹Artvin Çoruh Üniversitesi, Harita Mühendisliği Bölümü, 08100, Artvin/Türkiye.

²Hacettepe Üniversitesi, Geomatik Mühendisliği Bölümü, 06800, Ankara/Türkiye.

³Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, 34220, İstanbul/Türkiye.

⁴Yıldız Teknik Üniversitesi, Harita Mühendisliği Bölümü, 34220, İstanbul/Türkiye.

ARAŞTIRMA MAKALESİ

*Sorumlu yazar:

Mustafa Üstüner
mustuner@artvin.edu.tr

doi:

Yayın süreci

Geliş tarihi: 20.05.2020

Kabul tarihi: 13.08.2020

Basım tarihi: 30.09.2020

Özet

Son yıllarda, makine öğrenmesi ve veri bilimi alanındaki araştırmacılar özgün ve de özellikle topluluk öğrenmesi alanında yeni algoritmalar geliştirmeye başlamışlardır. Bu yeni nesil topluluk öğrenme algoritmalarından olan aşırı gradyan artırma (XGBoost) ve hafif gradyan artırma makineleri (LightGBM) yöntemleri, mevcut ve aynı zamanda sık kullanılan makine öğrenme algoritmalarına kıyasla daha yüksek performans gösterdiklerinden dolayı veri bilimindeki araştırmacıların ilgisini çekmiştir. Ancak uzaktan algılama görüntülerinin sınıflandırılması amacıyla henüz yeterli düzeyde test edilmemiştir. Bu çalışma kapsamında, XGBoost ve LightGBM algoritmalarının çok zamanlı polarimetrik sentetik açıklıklı radar (PolSAR) görüntüleri kullanılarak tarımsal ürünlerin sınıflandırılmasındaki performansları, hem işlem hızı hem de elde edilen sınıflandırma doğrulukları açısından karşılaştırılmıştır. Çalışma kapsamında, tam polarimetrik RADARSAT-2 uydu görüntülerine ait doğrusal geri saçılım değerleri kullanılmıştır. Tarımsal ürünlerin zamana bağlı dinamik olarak değişen yapılarından dolayı çalışmamızda çok zamanlı bir veri seti kullanılmıştır. Deneysel sonuçlarımız, LightGBM yönteminin hem işlem hızı hem de sınıflandırma performansı açısından XGBoost yöntemine göre daha üstün olduğunu göstermiştir, bu iki yöntem elde edilen doğruluklar sırası ile 0.860 ve 0.845'dir. Sınıflandırma sonuçlarının elde edilmesinde k-katlamalı (k=5) çapraz doğrulama tekniği kullanılmıştır.

Anahtar kelimeler: Uzaktan algılama, Sentetik açıklıklı radar, Hafif gradyan artırma makineleri

Abstract

In recent years, machine learning and data science communities have started to develop novel algorithms, especially in the area of ensemble learning. The new generation ensemble learning algorithms such as extreme gradient boosting (XGBoost) and light gradient boosting machines (LightGBM) have gained great attention in data science because of their greater performance compared to the state-of-art machine learning algorithms. However, they have not yet been fully tested for the classification of remotely-sensed images. This paper compares the performance of the XGBoost and LightGBM in terms of classification accuracy and computation time for crop classification using multi-temporal polarimetric SAR (PolSAR) data. The linear backscatter of full-polarimetric RADARSAT-2 were used as the polarimetric feature in this research. A multi-temporal dataset was used in our study because of the time-dynamic structure of crops. Our experimental results demonstrate that LightGBM yielded greater performance compared to XGBoost in terms of classification accuracy (0.860 vs 0.845) and computation cost. K-fold (k=5) cross validation was used to assess the classification results.

Keywords: Remote sensing, Synthetic aperture radar, Light gradient boosting machine

1. Giriş

Tarımsal ürünlerin sınıflandırılması, uzaktan algılama tabanlı tarım uygulamalarının başında gelmektedir ve çoğu uygulama için de önemli olan altlık veriler sağlamaktadır (Atzberger, 2013). Ürün rekoltesinin tahmin edilmesi ve tarımsal ürünlerin izlenmesinde, uzaktan algılama verileri sıklıkla kullanılmaktadır ve bu aşamada ürün türlerinin doğru tespit edilmesi büyük önem arz etmektedir (McNairn vd. 2002). Uzaktan algılama verileri, yer gözlem uydularının sinoptik görüş özelliği ve aynı bölgeyi farklı zaman aralıklarında tekrar görüntüleyebilme yeteneğinden dolayı tarım ile ilgili çalışmalarda, klasik yöntemlere nazaran, tercih edilmektedir (Villa vd. 2015; Khosravi ve Alavipanah, 2019). Tarımsal ürünlerin sınıflandırılması, arazi örtüsünün sınıflandırılması işlemine nazaran daha zorlu bir süreçtir. Bunun nedeni tarımsal ürünlerin yılın bazı periyotlarında birbirine yakın yansıma (reflektans) değerlerine sahip olmasıdır (Waldhoff vd. 2012; Hütt vd. 2016). Çok zamanlı (multi-temporal) veriler kullanıldığında ise, her bir tarım ürünü farklı fenolojik özellik gösterdiğinden tarımsal ürünlerin birbirlerinden ayırt edilmesi daha kolay olacaktır. Bu nedenle, tarımsal ürünlerin yüksek doğrulukla tespit edilmesinde çok zamanlı uydu görüntülerine ihtiyaç duyulmaktadır (Gómez-Chova vd. 2015; Hütt vd. 2016).

Optik uzaktan algılama görüntüleri (RapidEye, Sentinel-2, Landsat-8, Worldview-2 vb.) yüksek/orta çözünürlükte spektral ve mekânsal bilgi içerdiklerinden dolayı tarımsal ürünlerin sınıflandırılmasında başarılı sonuçlar vermektedir. Bu başarının altında yatan ana nedenler, tarımsal ürünlerin farklı spektral yansıma değerlerine sahip olmaları ve zaman bazlı analizlerde ürünlerin yansıma değerindeki değişimlerin (fenolojik değişimler) birbirinden farklı olmasıdır (Conrad vd. 2014; Inglada vd. 2015). Optik uzaktan algılama sistemleri pasif sistemlerdir ve görüntü alımında bazı kısıtlamalara sahiptir. Atmosferik koşullardan kaynaklanan hatalar (bulut ve sis) ve güneş enerjisine duyulan ihtiyaç nedeniyle bazı durumlarda istenilen kalitede görüntü elde edilememektedir. Özellikle katma değeri yüksek tarımsal ürünlerin belli zaman aralıklarında izlenmesinde, optik uydu görüntüleri yukarıda bahsedilen kısıtlamalardan kaynaklı yetersiz kalmaktadır. Bu tarz görüntü alımındaki kısıtlamaların önüne SAR (Sentetik Açıklıklı Radar) görüntüleri kullanılarak geçmek mümkündür. SAR görüntüleri, optik görüntülere nazaran bulut, sis vb. durumlardan etkilenmemektedir ve zamansal izlemenin kritik öneme sahip olduğu tarım projelerinde sıklıkla tercih edilmektedir. Bu avantajının yanı sıra, karmaşık SAR sinyali (geri saçılan sinyalin faz ve genlik bilgisinin birlikte kaydedilmesi nedeni ile) hedef objenin fiziksel ve geometrik yapısına karşı duyarlıdır ve ürünlerin birbirinden ayrılmasında önemli rol oynamaktadır (Jiao vd. 2014; McNairn ve Shang, 2016; Skakun vd. 2016). Bu nedenle, özellikle son yıllarda SAR görüntülerinin tarımsal ürünlerin sınıflandırılmasındaki kullanımı giderek artmaktadır. Diğer bir önemli etken ise, Avrupa Uzay Ajansı (ESA) kopernik programı kapsamında fırlatılan Sentinel-1 (A/B) SAR uydu görüntülerinin ücretsiz olarak elde edilebiliyor olmasıdır. Optik görüntüler ile kıyaslandığında, SAR ya da tam polarimetrik SAR (PolSAR) görüntülerinin tarımsal ürünlerin sınıflandırılmasındaki kullanımı daha az sayıdadır (Khosravi ve Alavipanah, 2019). Yapılan çalışmalar incelendiğinde, SAR/PolSAR görüntülerinin sınıflandırılmasında en sık kullanılan algoritmalar maksimum olabilirlik (maximum likelihood), yapay sinir ağları, karar ağaçları, rastgele orman (RO) ve destek vektör makineleri (DVM) yöntemleridir (Skakun vd. 2016). DVM ve RO algoritmaları parametrik olmayan yapıları ve üstün sınıflandırma kabiliyetleri nedeniyle diğer yöntemlere kıyasla uzaktan algılamada daha sık kullanılmıştır (Pal, 2012; Petropoulos vd. 2012). İki sınıfı birbirinden ayıran en uygun hiperdüzlemin belirlenmesi esasına dayanan DVM, sınıfların doğrusal düzlemlerle ayrılmasını durumunda çekirdek fonksiyonlarından faydalanmaktadır. Bu noktada, çekirdek türünün seçimi ve çekirdeğe ait en uygun parametrelerin belirlenmesi önem arz etmektedir (Melgani ve Bruzzone, 2004; Kavzoglu ve Colkesen, 2009). Karar ağaçları tabanlı bir yöntem olan RO algoritması ise, destek vektör makinelerine kıyasla daha az sayıda parametreye ihtiyaç duyması ve daha hızlı olması nedeniyle tercih edilmektedir (Belgiu ve Drăguț, 2016).

Rastgele orman algoritması uzaktan algılama görüntülerinin sınıflandırılmasında kullanılan en yaygın topluluk öğrenme algoritmasıdır. Diğer yaygın kullanılan topluluk öğrenme algoritmaları ise torbalama (bagging) ve hızlandırma (boosting) algoritmalarıdır (Akar ve Güngör, 2012; Belgiu ve Drăguț, 2016). Örüntü tanıma ve makine öğrenmesi alanlarındaki son gelişmeler ile birlikte, yeni nesil topluluk öğrenme algoritmaları ortaya çıkmış ve başarılı bir şekilde kullanılmaya başlanmıştır. Uzaktan algılama alanında da özellikle son yıllarda kullanılmaya başlanan yeni nesil topluluk öğrenme algoritmalarına aşırı gradyan artırma (extreme gradient boosting, XGBoost) ve hafif gradyan artırma makineleri (light gradient boosting machines, LightGBM) yöntemleri örnek olarak verilebilir (Georganos vd. 2018; Üstüner ve Balık Şanlı, 2019; Abdi, 2020). Bu algoritmaları uzaktan algılama görüntülerinin sınıflandırılması amacıyla kullanan çalışmalar sınırlı sayıdadır. Georganos vd. (2018) yüksek çözünürlüklü uydu görüntülerinden faydalanarak arazi örtüsü ve kullanımının sınıflandırılmasında XGBoost, DVM, RO, k-en yakın komşuluk ve özyinelemeli bölümlenme (recursive partitioning) yöntemlerini kullanmıştır. Shi vd. (2019) çalışmalarında, LiDAR nokta bulutu verisinin sınıflandırılması işleminde LightGBM yönteminden faydalanmışlardır. Jiang vd. (2019) şeker kamışının tespit edilmesi işleminde XGBoost ve RO algoritmalarının performansını karşılaştırmıştır. Moorthy vd. (2020) ise, LiDAR nokta bulutu verisinden yaprak ve ağaçlık alanların sınıflandırılması işleminde RO, XGBoost ve LightGBM yöntemlerinin performanslarını karşılaştırmıştır.

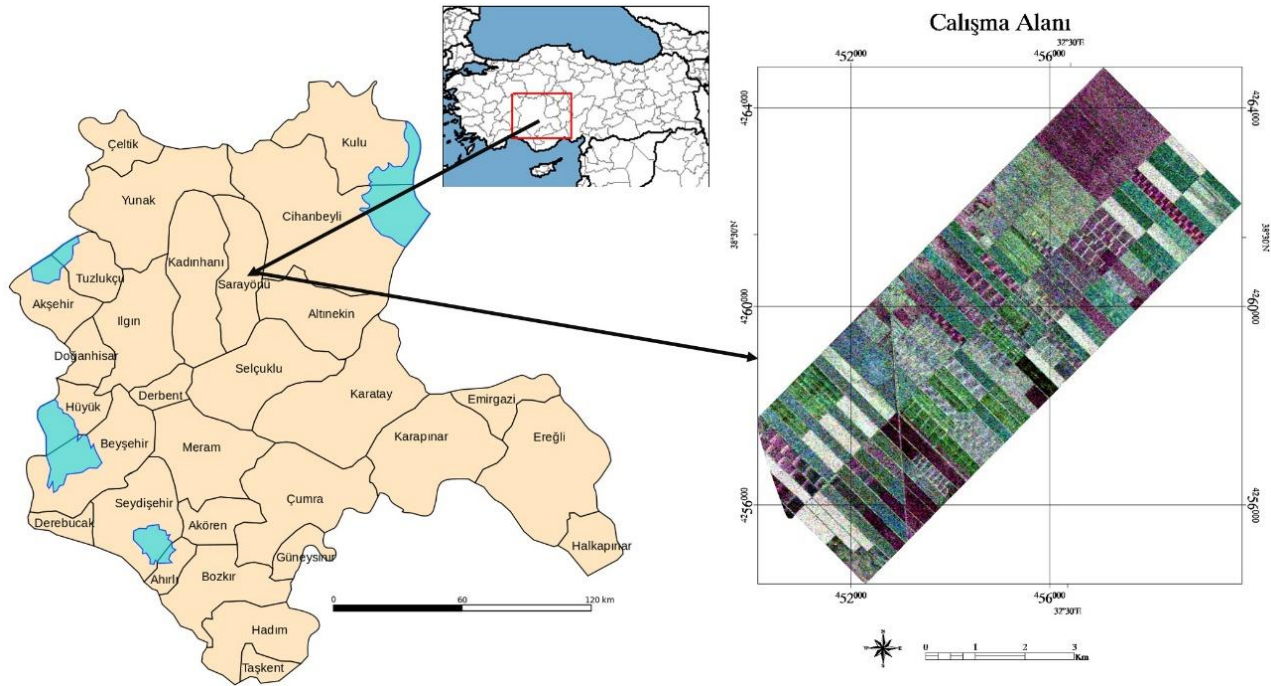
Tarım ürünlerinin uzaktan algılama görüntülerinden sınıflandırılmasında ise, en sık kullanılan algoritmalar maksimum olabilirlik, RO, DVM, derin öğrenme ve XGBoost yöntemleridir. Zhong vd. (2019) çalışmalarında, tarım ürünlerinin sınıflandırılması amacıyla derin öğrenme, XGBoost, DVM ve RO yöntemlerinden faydalanmışlardır. Saini ve Ghosh (2019) çalışmalarında Sentinel-2 görüntülerinden faydalanmış ve tarımsal ürünlerin sınıflandırılması işleminde XGBoost, RO ve DVM algoritmalarının performansını karşılaştırmıştır. Dey vd. (2020) ise çalışmalarında, tam polarimetrik Radarsat-2 PolSAR uydu görüntülerinden faydalanılmış ve çalışmalarında XGBoost ve RO algoritmalarının tarımsal ürünlerin sınıflandırılmasındaki performansını karşılaştırmıştır.

Yukarıda listelenen çalışmalardan sadece Moorthy vd. (2020) LightGBM ve XGBoost yöntemlerini karşılaştırmıştır fakat o çalışma ise tarımsal ürünlerin sınıflandırılması ile ilişkili değildir. Bu iki yöntemin (LightGBM ve XGBoost yöntemleri) tarımsal ürünlerin sınıflandırılmasındaki karşılaştırmalı analizini yapan bir çalışmaya henüz literatürde rastlanmamıştır.

Bu çalışmada, çok zamanlı PolSAR uydu görüntülerinin tarımsal ürünlerin sınıflandırılmasındaki kullanımı XGBoost ve LightGBM yöntemleri ile araştırılmıştır. Kullanılan yöntemlerin hem işlem hızı hem de sınıflandırma performansı açısından karşılaştırmalı analizi yapılmıştır. Sınıflandırma doğruluklarının (başarımlarının) belirlenmesinde ise k-katlamalı çapraz doğrulama yöntemi kullanılmıştır.

2. Çalışma Alanı ve Kullanılan Veriler

Konya ilinin Sarayönü ilçesine bağlı Gözlü köyü sınırları içerisinde yer alan çalışma alanı, Konya şehir merkezinin 65 km kuzeyinde yer almakta ve yaklaşık olarak 42 km²lik bir alanı kapsamaktadır (Şekil 1). Tarımsal üretim ve hayvancılık faaliyetlerinin sürdürüldüğü ilçede buğday, arpa, mısır, patates ve ayçiçeği üretimleri yapılmaktadır. Geniş düzlüklere sahip olan Konya ovası, ülke ölçeğinde tarımsal üretimin yapılmasında önemli bir yere sahiptir. Tamamen tarım ürünleri ile kaplı olan çalışma alanında, yem bitkisi, mısır, patates, ayçiçeği ve buğday olmak üzere beş farklı ürün türü bulunmaktadır. Tarımsal ürünlerin zamana bağlı dinamik olarak değişen yapılarından dolayı çalışmamızda çok zamanlı bir veri seti kullanılmıştır.



Şekil 1. Çalışma Alanı

Çalışma kapsamında, tek bakışlı karmaşık (single look complex) görüntü formatındaki tam polarimetrik Radarsat-2 PolSAR uydu görüntülerinden faydalanılmıştır. Kullanılan PolSAR görüntüsünün özellikleri ve görüntülerin alım tarihleri Tablo 1'de görülmektedir.

Tablo 1. Radarsat-2 görüntüsünün özellikleri

Algılayıcı	Radarsat-2
Dalga Boyu	C-Bant (5.6 cm)
Çözünürlük	4.7 m x 5.1 m (Menzil x Azimut)
Geliş Açısı	40°
Polarizasyon (Alım Modu)	Tam Polarimetrik
Görüntü Alım Tarihleri	13.06.2016
	07.07.2016
	31.07.2016
	24.08.2016

Çalışma kapsamında, uydu görüntülerinin alım tarihleri ile eş zamanlı arazi çalışması gerçekleştirilmiş ve ürünlere ilişkin örnek noktalar el GPS'i yardımıyla toplanmıştır. Arazi çalışmasının ardından, örnek noktalardan faydalanarak tarla bazında (poligon olarak) yer gerçekliği verisi oluşturulmuş ve bu veriler hem sınıflandırma hem de doğruluk analizi işlemlerinde kullanılmıştır (Tablo 2).

Tablo 2. Yer gerçekliği verisi

Sınıf	Yer Gerçekliği Verisi (Piksel Sayısı)
Yem Bitkisi	5460
Mısır	5581
Patates	6093
Ayçiçeği	5361
Buğday	5481

Tek bakışlı karmaşık görüntü formatında elde edilen görüntülerden, doğrusal geri saçılım değerlerini içeren polarimetrik özelliklerin oluşturulması için bazı ön işleme adımları gerekmektedir. Bu ön işleme adımları Avrupa Uzay Ajansı'nın ücretsiz olarak sağladığı SNAP (Sentinel Application Platform) yazılımı (versiyon 6.0) ile gerçekleştirilmiştir. Bu ön işleme adımları sırasıyla, görüntünün radyometrik (sigma nought) kalibrasyonu, benek (speckle) filtreleme, ortorektifikasyon ve piksel değerlerinin radar geri saçılım (dB) değerine dönüştürülmesi işlemleridir. Bu temel ön işleme adımları dört farklı tarih için de uygulanıp çok zamanlı bir veriseti (time-stack) oluşturulmuştur.

3. Sınıflandırma Yöntemleri

3.1 Aşırı Gradyan Artırma (XGBoost)

Aşırı Gradyan Artırma (XGBoost), yüksek sınıflandırma performansı elde etmesi ile hem veri bilimi hem de uzaktan algılama alanında gittikçe popüler hale gelen ve gradyanı artırılmış (gradient boosted) karar ağaçları temeline dayalı çalışan bir makine öğrenme algoritmasıdır (Chen ve Guestrin, 2016; Georganos vd. 2018). Yeni nesil topluluk öğrenme algoritmalarından olan XGBoost, algoritmanın eğitilmesi sürecinde aşırı uyumlama (overfitting) sorununun önüne geçerek modelin genel doğruluğunu (performansını) artırmaktadır. Bu yöntemin başarısının altında yatan ana neden, öğrenme sürecinde kullandığı amaç fonksiyonudur. Amaç fonksiyonu, kayıp/yitim fonksiyonu (Loss function) ve düzenleştirme (regularization) teriminden oluşmaktadır. Kayıp/yitim fonksiyonu, modelin yaptığı her bir tahmin değerinin (predicted class) gerçek değerinden olan farkını hesaplamaktadır. Düzenleştirme terimi ise, modelin karmaşıklığını kontrol etmekte ve bu da modeldeki aşırı uyumlama sorununu ortadan kaldırmaktadır (Chen ve Guestrin, 2016; Rumora vd. 2020; Abdi, 2020). Bu çalışma kapsamında amaç fonksiyonu olarak, softmax çok sınıflı amaç fonksiyonu kullanılmıştır. XGBoost sınıflandırma işlemi, açık kaynaklı Python XGBoost paketinden (versiyon 1.0.2) faydalanarak gerçekleştirilmiştir (XGBoost, 2020) ve çalışma kapsamında belirlenen parametreler Tablo 3'te verilmiştir.

En uygun parametre değerleri (yaprak sayısı ve öğrenme oranı parametreleri için) ızgara arama (grid search) algoritması ile bulunmuştur. En uygun parametreler, yaprak sayısı için [100-1000] ve öğrenme oranı için ise [0.1-1] değerleri arasında aratılmıştır. Izgara arama algoritması, parametre çifti (yaprak sayısı ve öğrenme oranı parametreleri) için tüm kombinasyonları eğitir ve çapraz doğrulama tekniği ile model doğruluğunu hesaplar. En yüksek model doğruluğunu veren parametre çifti, en uygun parametre değeri olarak kabul edilir (Ayhan ve Erdoğan, 2014; Maxwell vd. 2018).

Tablo 3. XGBoost parametreleri

Parametre	Değer
Hızlandırma Türü	DART (Dropouts meet Multiple Additive Regression Trees)
Maksimum Derinlik	5
Özellik Alt-Örnekleme	1
Yaprak Sayısı	100
Öğrenme Oranı	0.1

3.2 Hafif Gradyan Artırma Makineleri

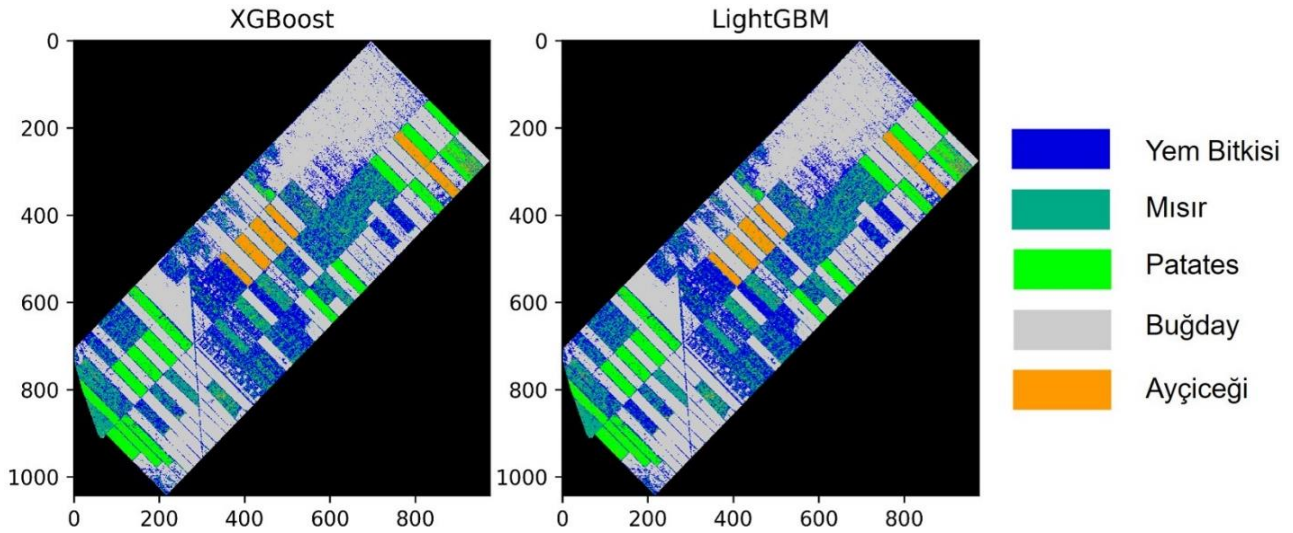
Hafif gradyan artırma makineleri (LightGBM), gradyan artırma çatısı altında çalışan karar ağaçları tabanlı yeni nesil topluluk öğrenme algoritmalarındandır. 2017 yılında Microsoft tarafından geliştirilen LightGBM, özellikle son yıllarda makine öğrenmesi yarışmalarında gösterdiği üstün başarıları ile hem veri bilimi hem de uzaktan algılama alanındaki araştırmacıların dikkatini çekmiştir (Ke vd. 2017; Liu vd. 2017; Ustuner ve Balık Sanli, 2019; Moorthy vd. 2020). Yöntemin isminden de anlaşılacağı üzere yüksek işlem hızına sahip bir algoritma olduğu için "Light" ön takısını almıştır. Bu yöntemi, diğer gradyan artırma algoritmalarından ayıran bir özelliği ise karar ağaçlarının eğitilmesi aşamasında kullandığı büyüme stratejisidir. LightGBM düşey yönde büyüme stratejisini (yapraksal büyüme) kullanırken, diğer gradyan artırma yöntemleri ise seviyesel (level-wise) büyüme stratejisini kullanmaktadır (Ke vd. 2017; Li vd. 2019; Ustuner ve Balık Sanli, 2019). LightGBM'i ayrıcalıklı ve özgün kılan diğer önemli bir detay ise, bünyesinde barındırdığı ve işlem hızını artırmasını sağlayan iki yeni algoritmadır. Bu algoritmalar, gradyan tabanlı tek taraflı örnekleme (Gradient-based One-Side Sampling, GOSS) ve ayrıcalıklı öznelik desteleme (Exclusive Feature Bundling, EFB) yöntemleridir. Gradyan tabanlı tek taraflı örnekleme ile verinin tamamını kullanmak yerine veriden ürettiği alt örneklenmiş (subsampling) veri kümesini kullanır. Ayrıcalıklı özellik desteleme ile de seyrek yapıdaki özellikleri daha sık/yoğun özelliklere dönüştürerek işlem karmaşıklığını azaltmaktadır (Ke vd. 2017; Li vd. 2019; Ustuner ve Balık Sanli, 2019). LightGBM sınıflandırma işlemi için, Microsoft tarafından ücretsiz olarak erişilebilen açık kaynaklı Python lightgbm paketi (versiyon 2.3.1) kullanılmıştır (LightGBM, 2020). Sınıflandırma işleminde kullanılan parametreler aşağıdaki tabloda görülmektedir. Parametrelerin seçimi LightGBM'in ana sayfasındaki "parametre ayarlama" sayfasındaki öneriler dikkate alınarak seçilmiştir.

Tablo 4. LightGBM Parametreleri

Parametre	Değer
Hızlandırma Türü	GOSS
Maksimum Derinlik	5
Yaprak Sayısı	100
Öğrenme Oranı	0.1
Amaç Fonksiyonu	Çok-sınıflı
Özellik Alt-Örnekleme	1.0
Yapraktaki Minimum Veri Sayısı	20
İterasyon Sayısı	100

4. Uygulama

Sınıflandırılmış görüntülerin doğruluk analizinde k-katlamalı çapraz doğrulama tekniği kullanılmıştır. Bu yöntemde, k sayıda parçaya bölünen orijinal veri (yer gerçekliği verisi) setinin (k-1) sayıdaki parçası modeli eğitmek için kullanılırken, geriye kalan tek parça ise modelin doğruluğunu test etmek amacıyla kullanılır. Bu işlem k defa tekrar edilir ve her bir seferde farklı bir parça test/eğitim verisi olarak kullanılır. Model doğruluğu ise, elde edilen doğrulukların ortalamasına eşittir (Kohavi, 1995; Toosi vd., 2019). Çalışmamızda k değeri 5 olarak seçilmiştir. Sınıflandırılmış görüntüler Şekil 2'de sunulmuştur.



Şekil 2. Sınıflandırılmış Görüntüler

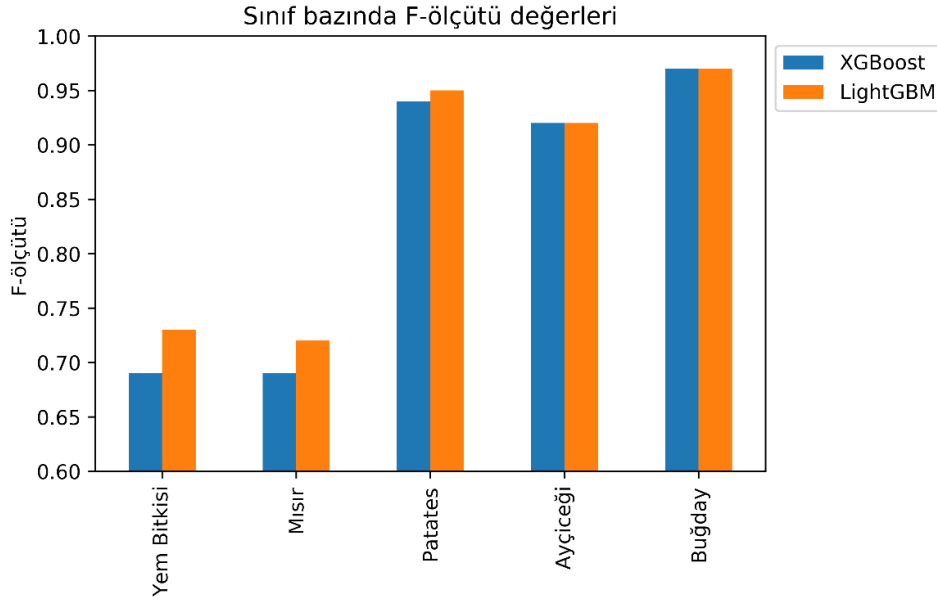
Çok zamanlı bir veri setinin kullanıldığı çalışmamızda, LightGBM yönteminin, XGBoost yöntemine göre hem işlem hızı (7.11 sn karşın 49.25 sn) hem de sınıflandırma performansı (0.860 karşın 0.845) açısından daha üstün olduğu sonucuna varılmıştır. Verilerin sınıflandırılması ve doğruluk analizi işlemleri Intel Core i7 işlemci ve 8GB belleğe (RAM) sahip bir bilgisayarda Windows 10 işletim sistemi altında gerçekleştirilmiştir. Sınıflandırma ve doğruluk analizi işlemlerinde Python açık kaynak kodlu XGBoost (versiyon 1.0.2), LightGBM (versiyon 2.3.1) ve Scikit-learn (versiyon 0.19) (Pedregosa vd. 2011) kütüphaneleri kullanılmıştır.

Tablo 5. Doğruluk değerleri ve işlem süreleri

Yöntem	Doğruluk (+/- Standart Sapma)	İşlem Süresi (sn)
XGBoost	0.845 (+/- 0.035)	49.25
LightGBM	0.860 (+/- 0.027)	7.11

Sınıflandırma doğrulukları arasındaki farkın istatistiksel olarak anlamlı olup olmadığı McNemar's testi kullanılarak analiz edilmiştir (Abdikan vd. 2015). Ki-kare dağılımını esas alan McNemar's testine göre, çalışmamızda χ^2 test istatistiği %95 güven aralığındaki kritik değerden ($\chi^2=3.84$) büyük olduğundan sınıflandırma sonuçları arasındaki farkın anlamlı olduğu sonucuna varılmıştır. Sınıf bazındaki doğrulukların analiz edilmesinde ise, kullanıcı doğruluğu (KD) ve üretici doğruluğu (ÜD) değerlerinin harmonik ortalamasına eşit olan F-ölçütü (F-score) değerleri kullanılmıştır. Sınıf bazında F-ölçütü değerleri Şekil 3'te görülmektedir.

$$F - \text{ölçütü} = 2 \times \frac{(\text{ÜD} \times \text{KD})}{(\text{ÜD} + \text{KD})} \quad (1)$$



Şekil 3. Sınıf bazında F-ölçütü değerleri

Ayçiçeği ve buğday sınıfları için, her iki yöntem de eşit F-ölçütü değerleri elde etmiştir fakat diğer üç sınıf türü (yem bitkisi, mısır ve patates) için LightGBM daha yüksek değerler elde etmiştir. Bu farklılığın ana nedeni, aynı eğitim verisi kullanılmasına rağmen algoritmaların öğrenme sürecindeki yapılarının birbirinden farklı olmasıdır. Patates, ayçiçeği ve buğday sınıfları için F-ölçütü değerleri 0.90'nın üzerinde elde edilirken, mısır ve yem bitkisi için bu değer 0.75 değerinin altında kalmıştır. Sınıf bazında F-ölçütü değerleri incelendiğinde ise, en yüksek değer (0.97) buğday sınıfına ait olurken, en düşük değer (0.69) ise XGBoost yöntemi ile yem bitkisi sınıfına aittir.

5. Sonuç ve Öneriler

Bu çalışmada, yeni nesil topluluk öğrenme algoritmalarından XGBoost ve LightGBM yöntemlerinin tarımsal ürünlerin sınıflandırılmasındaki performansları karşılaştırılmıştır. Bu performans karşılaştırması, hem model doğruluğu hem de işlem süresi baz alınarak yapılmıştır. Deneysel sonuçlarımız, tarımsal ürünlerin sınıflandırılması işleminde LightGBM yönteminin XGBoost yöntemine kıyasla daha üstün olduğu sonucuna varmıştır. LightGBM yönteminin işlem süresi olarak XGBoost yönteminden yedi kat daha hızlı olduğu tespit edilmiştir. Fakat iki yöntem için de en uygun parametrelerin bulunması hem yorucu hem de vakit alan bir süreçtir. Kullanılan girdi verinin boyutu ve sınıf türlerinin sayısına/karmaşıklığına bağlı olarak değişkenlik gösterebilen en uygun parametreler, bu iki yöntem için de üzerinde durulması gereken önemli konulardır. Bu çalışma ile ayrıca, çok zamanlı PolSAR verisinin tarımsal ürünlerin sınıflandırılmasındaki kullanılabilirliği test edilmiştir.

Gelecek çalışmalarımızda, daha çok sayıda sınıfa sahip verilerle bu kıyaslamaların tekrar yapılması ve özellikle kullanılan parametrelerin sınıflandırma doğruluğuna olan etkilerinin araştırılması hedeflenmektedir. PolSAR verisi kullanılarak yapılan bu çalışmanın, optik veriler üzerinde de test edilmesi planlanmaktadır.

Teşekkür

Bu çalışma, Yıldız Teknik Üniversitesi (YTÜ) Bilimsel Araştırma Projeleri Koordinatörlüğü tarafından FBA-2017-3062 kodlu proje kapsamında desteklenmiştir. Destekleri için Yıldız Teknik Üniversitesi (YTÜ) Bilimsel Araştırma Projeleri Koordinatörlüğü'ne teşekkür ederiz.

Kaynaklar

- Abdi, A. M. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sensing*, 57(1), 1-20, doi: 10.1080/15481603.2019.1650447.
- Abdikan, S., Bilgin, G., Sanlı, F. B., Uslu, E., & Ustuner, M. (2015). Enhancing land use classification with fusing dual-polarized TerraSAR-X and multispectral RapidEye data. *Journal of Applied Remote Sensing*, 9(1), 096054, doi: 10.1117/1.JRS.9.096054.

- Akar, Ö., & Güngör, O. (2012). Classification of multispectral images using Random Forest algorithm. *Journal of Geodesy and Geoinformation*, 1(2), 105-112.
- Atzberger, C. (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sensing*, 5(2), 949-981.
- Ayhan, S., & Erdoğan, Ş. (2014). Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 9(1), 175-201.
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016. Proceedings*. (pp. 785–794).
- Conrad, C., Dech, S., Dubovyk, O., Fritsch, S., Klein, D., Löw, F., ... & Zeidler, J. (2014). Derivation of temporal windows for accurate crop discrimination in heterogeneous croplands of Uzbekistan using multitemporal RapidEye images. *Computers and Electronics in Agriculture*, 103, 63-74.
- Dey, S., Mandal, D., Robertson, L. D., Banerjee, B., Kumar, V., McNairn, H., Bhattacharya, A. & Rao, Y. S. (2020). In-season crop classification using elements of the Kennaugh matrix derived from polarimetric RADARSAT-2 SAR data. *International Journal of Applied Earth Observation and Geoinformation*, 88, 102059, doi: 10.1016/j.jag.2020.102059.
- Georganos, S., Grippa, T., Vanhuyse, S., Lennert, M., Shimoni, M., Kalogirou, S., & Wolff, E. (2018). Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. *GIScience & Remote Sensing*, 55(2), 221-242.
- Gómez-Chova, L., Tuia, D., Moser, G., & Camps-Valls, G. (2015). Multimodal classification of remote sensing images: A review and future directions. In *Proceedings of the IEEE*, 103(9), (pp. 1560-1584). IEEE.
- Hütt, C., Koppe, W., Miao, Y., & Bareth, G. (2016). Best accuracy land use/land cover (LULC) classification to derive crop types using multitemporal, multisensor, and multi-polarization SAR satellite images. *Remote Sensing*, 8(8), 684, doi: 10.3390/rs8080684.
- Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., ... & Koetz, B. (2015). Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, 7(9), 12356-12379.
- Jiang, H., Li, D., Jing, W., Xu, J., Huang, J., Yang, J., & Chen, S. (2019). Early season mapping of sugarcane by applying machine learning algorithms to Sentinel-1A/2 time series data: a case study in Zhanjiang City, China. *Remote Sensing*, 11(7), 861, doi: 10.3390/rs11070861.
- Jiao, X., Kovacs, J. M., Shang, J., McNairn, H., Walters, D., Ma, B., & Geng, X. (2014). Object-oriented crop mapping and monitoring using multi-temporal polarimetric RADARSAT-2 data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 96, 38-46.
- 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.
- Khosravi, I., & Alavipanah, S. K. (2019). A random forest-based framework for crop mapping using temporal, spectral, textural and polarimetric observations. *International Journal of Remote Sensing*, 40(18), 7221-7251.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems (NIPS 2017), 2017. Proceedings*. (pp. 3146-3154).
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *14th international joint conference on Artificial intelligence (IJCAI'95), 1995. Proceedings*. (pp. 1137-1145).
- Li, W., Ding, S., Chen, Y., Wang, H., & Yang, S. (2019). Transfer learning-based default prediction model for consumer credit in China. *The Journal of Supercomputing*, 75(2), 862-884.
- LightGBM, (2020, June 15). *LightGBM Python API*, Retrieved from <https://lightgbm.readthedocs.io/en/latest/Python-API.html>.
- Liu, L., Ji, M., & Buchroithner, M. (2017). Combining partial least squares and the gradient-boosting method for soil property retrieval using visible near-infrared shortwave infrared spectra. *Remote Sensing*, 9(12), 1299, doi: 10.3390/rs9121299.
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), 2784-2817.
- McNairn, H., & Shang, J. (2016). A review of multitemporal synthetic aperture radar (SAR) for crop monitoring. In Y. Ban (Eds.), *Multitemporal Remote Sensing: Methods and Applications*, (pp. 317-340), Cham, Switzerland: Springer International Publishing AG.
- McNairn, H., Ellis, J., Van Der Sanden, J. J., Hirose, T., & Brown, R. J. (2002). Providing crop information using RADARSAT-1 and satellite optical imagery. *International Journal of Remote Sensing*, 23(5), 851-870.

- Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8), 1778-1790.
- Moorthy, S. M. K., Calders, K., Vicari, M. B., & Verbeeck, H. (2019). Improved supervised learning-based approach for leaf and wood classification from LiDAR point clouds of forests. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5), 3057-3070.
- Pal, M. (2012). Advanced algorithms for land use and cover classification. In X. Yang & J. Li (Eds.), *Advances in Mapping from Remote Sensor Imagery: Techniques and Applications*, (pp. 70-82), Boca Raton: CRC Press.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Petropoulos, G. P., Kalaitzidis, C., & Vadrevu, K. P. (2012). Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Computers & Geosciences*, 41, 99-107.
- Rumora, L., Miler, M., & Medak, D. (2020). Impact of Various Atmospheric Corrections on Sentinel-2 Land Cover Classification Accuracy Using Machine Learning Classifiers. *ISPRS International Journal of Geo-Information*, 9(4), 277, doi: 10.3390/ijgi9040277.
- Saini, R., & Ghosh, S. K. (2019). Crop classification in a heterogeneous agricultural environment using ensemble classifiers and single-date Sentinel-2A imagery. *Geocarto International*, 1-19, doi: 10.1080/10106049.2019.1700556.
- Shi, X., Cheng, Y., & Xue, D. (2019, October). Classification Algorithm of Urban Point Cloud Data based on LightGBM. In *IOP Conference Series: Materials Science and Engineering* (Vol. 631, No. 5, p. 052041). IOP Publishing.
- Skakun, S., Kussul, N., Shelestov, A. Y., Lavreniuk, M., & Kussul, O. (2016). Efficiency assessment of multitemporal C-band Radarsat-2 intensity and Landsat-8 surface reflectance satellite imagery for crop classification in Ukraine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8), 3712-3719.
- Toosi, N. B., Soffianian, A. R., Fakheran, S., Pourmanafi, S., Ginzler, C., & Waser, L. T. (2019). Comparing different classification algorithms for monitoring mangrove cover changes in southern Iran. *Global Ecology and Conservation*, 19, e00662, doi: 10.1016/j.gecco.2019.e00662.
- Ustuner, M., & Balık Sanli, F. (2019). Polarimetric Target Decompositions and Light Gradient Boosting Machine for Crop Classification: A Comparative Evaluation. *ISPRS International Journal of Geo-Information*, 8(2), 97, doi: 10.3390/ijgi8020097.
- Villa, P., Stroppiana, D., Fontanelli, G., Azar, R., & Brivio, P. A. (2015). In-season mapping of crop type with optical and X-band SAR data: A classification tree approach using synoptic seasonal features. *Remote Sensing*, 7(10), 12859-12886.
- Waldhoff, G., Curdt, C., Hoffmeister, D., & Bareth, G. (2012). Analysis of multitemporal and multisensor remote sensing data for crop rotation mapping. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, 177-182.
- XGBoost, (2020, June 15). *XGBoost Python Package*, Retrieved from <https://xgboost.readthedocs.io/en/latest/python/index.html>.
- Zhong, L., Hu, L., & Zhou, H. (2019). Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, 221, 430-443.