



Araştırma Makalesi/Research Article

## Sentinel 2 Uydu Görüntülerinden Bazı Bitki Türlerinin Makine Öğrenmesi ile Belirlenmesi

Emre Tunca\* 

Eyüp Selim Köksal 

<sup>1</sup>Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Tarımsal Yapılar ve Sulama Bölümü, Samsun  
\*Sorumlu yazar: emre.tunca@omu.edu.tr

Geliş Tarihi: 17.12.2020

Kabul Tarihi: 03.03.2021

### Öz

Uydu görüntülerinden bitkisel ürünlerin sınıflandırılması tarım alanlarının yönetimi, gıda güvenliğinin sağlanması ve tarımsal politikaların oluşturulması için oldukça önemli bilgiler sağlar. Bitki türleri genel olarak uydu görüntülerinden hesaplanan vejetasyon indekslerine dayalı olarak veya çeşitli görüntü sınıflandırma teknikleri ile tahmin edilmektedir. Fakat bu yaklaşımlarda farklı bitkilerin benzer fenolojik ve spektral özelliklere sahip olması nedeniyle başarı oranı düşüktür. Bu nedenle bitki türlerinin uydu görüntüleri ile sınıflandırılması işleminde yeni, hassas ve başarılı bir yaklaşıma ihtiyaç duyulmaktadır. Bu çalışmanın amacı Rassal Orman (RO), Destek Vektör Makinesi (DVM) ve K-En Yakın Komşu (K-NN) makine öğrenme algoritmaları kullanarak uydu görüntülerinden bitki türlerinin sınıflandırılmasıdır. Çalışma kapsamında 2020 yılı Gökhöyük Tarım İşletmesi Müdürlüğü'ne ait bitkisel üretim sahasında yetiştirilen bitkilerin sınıflandırılmasında zaman serisi biçiminde Sentinel 2 uydu görüntüleri kullanılmıştır. Çalışmadan elde edilen sonuçlara göre en başarılı sınıflandırma (%95.3) RO ile hesaplanırken en düşük başarı DVM ile elde edilmiştir (%75.9). K-NN ile yapılan sınıflandırma başarısı ise %91.8 olarak hesaplanmıştır.

**Anahtar Kelimeler:** makine öğrenmesi, rassal orman, destek vektör makinesi, k-en yakın komşu, bitki sınıflandırması, sentinel 2

### Determination of Some Plant Species from Sentinel 2 Satellite Images by Machine Learning Abstract

Crop classification provides important information for agricultural management, ensuring food security and developing agricultural policy. Generally, crop species are estimated based on vegetation indexes calculated from satellite images or by using various image classification techniques. However, this approach has low success rate due to different crop species have similar phenological and spectral features. For this reason, a new, accurate and robust approach is needed. The aim of this study was to classify crop from satellite image by using Random Forest (RO), Support Vector Machines (DVM) and K-Nearest Neighbors (K-NN) machine learning algorithms. Within the scope of the study, Sentinel 2 satellite images were used in time series format for the crop classification of grown in the field of Gökhöyük Agricultural Enterprise Directorate in 2020. According to the result of this study, while the most successful classification result obtained from RO, the lowest success was obtained from DVM. Classification success with K-NN was calculated as 91.8%.

**Keywords:** machine learning, random forest, support vector machine, k-nearest neighbor, crop classification, sentinel 2

### Giriş

Tarım sektörü, bir ülkedeki en önemli ekonomi kaynaklarından birisidir. Tarımsal üretimin artırılması ekonomik gelişmenin temel anahtarları arasındadır (Baker, 1987). Gelişmiş veya gelişmekte olan ülkelerde tarım ve sanayi sektörünün birbiri ile uyumlu olarak gelişmesi gerekmektedir. Bu nedenle tarım, ülke ekonomilerindeki stratejik sektörlerin başında gelmektedir (Doğan ve ark., 2015). Fakat son yıllarda ortaya çıkan nüfus artışı ile birlikte gıda güvenliği sorunu ve beslenme alışkanlıklarındaki değişiklikler gıda savaşları beklentisini ortaya çıkartmıştır (Uzundumlu, 2012). Bu sebeple tarım sektöründeki politikalar oldukça önem taşımaktadır. Bu politikaların belirlenmesinde ise temel dayanak tarımsal üretime ait veri ve bilgilerdir. Söz konusu bilgilerin hassas ve detaylı bir biçimde kayıt altına alınması tarımda karar vericiler için oldukça önemlidir.



Tarımsal üretime ait bilgiler genel olarak çiftçi beyanlarına dayalı olarak oluşturulurken, araziye ait veriler ise kadastro kayıtları doğrultusunda oluşturulmaktadır. Geleneksel olarak gerçekleştirilen bu yöntemde tapu bilgileri ile arazi kullanımı arasındaki uyumsuzluklar ve çiftçilerin yanlış veya eksik beyanda bulunabilmesi sebebiyle tarımsal istatistikler ile arazi gerçekleri arasında birtakım farklılıklar meydana gelebilmektedir. Bu farklılığın ortaya koyulmasında tercih edilebilecek yöntemlerden birisi arazi ziyaretidir ve teknik bir personel tarafından yapılması gereken bu kontrollerinin maliyeti oldukça yüksektir.

Son 20 yıldır, uzaktan algılanmış görüntüler ile çeşitli sınıflandırma yöntemleri kullanılarak bitki türü ve bitkisel üretim alanları başarılı bir biçimde tahmin edilmektedir (Li ve ark., 2020). Bu amaçla En Yüksek Benzerlik (MLC) görüntü sınıflandırma tekniği en sık tercih edilen yöntem olarak öne çıkmaktadır (Maxwell ve ark., 2018). Fakat MLC, verinin normal bir dağılım gösterdiğini varsayan parametrik bir yöntemdir (Ok ve ark., 2012). Bu nedenle bitki spektral karakteristiğine dayalı gerçekleştirilen sınıflandırma başarısı istenilen seviyede değildir. Son yıllarda makine öğrenmesi algoritmalarında önemli gelişmeler meydana gelmiştir. Söz konusu algoritmaların iyileştirilmesi ve bilgisayar donanımlarındaki gelişmelerle birlikte makine öğrenmesi algoritmaları tarım sektöründe uygulanmaya başlanmıştır. Örneğin, 2001 yılında ABD’deki arazi kullanım veri tabanı (NLCD) karar ağacı algoritması ile başarılı bir biçimde oluşturulmuştur (Homer ve ark., 2004). Hooda ve ark. (2006) tarafından gerçekleştirilen bir çalışmada Hindistan’da buğday üretim alanlarının uydu görüntülerinden makine öğrenmesi ile tespit edilmesi amaçlanmıştır. Çalışma sonucunda %5 hata ile buğday üretim alanlarının tespit edildiği bildirilmiştir. Yang ve ark. (2011)’de multispektral veri ve makine öğrenmesi algoritmaları kullanılarak bitki türlerinin başarılı bir biçimde ayrıştırılabileceği bildirilmiştir. Jay ve ark. (2009)’da uzaktan algılanmış uydu görüntüsündeki çeşitli bitki gruplarının sınıflandırması Rastal Orman (RO) algoritması kullanılarak %88.37 başarı ile gerçekleştirilmiştir. Waske ve Braun (2009) MLC sınıflandırma yöntemi ile RO makine öğrenme algoritmasının bitki türü sınıflandırılmasındaki başarısını karşılaştırmıştır. Çalışma sonucunda RO’nun MLC’den yaklaşık %10 daha başarılı sınıflandırma gerçekleştirdiği belirtilmiştir. Yapılan bir diğer çalışmada eğitimli sınıflandırma yöntemleri (en düşük mesafe, paralel bağlantı ve MLC) ile K-En yakın Komşu (K-NN) makine öğrenmesi algoritmasının uzaktan algılanmış görüntülerde bitki sınıflandırma başarısı karşılaştırılmıştır. Çalışma sonucunda K-NN algoritmasının diğer yöntemlerden daha başarılı olduğu bildirilmiştir (Prasad ve ark., 2006). RO ve K-NN algoritmalarına ek olarak Destek Vektör Makinesi (DVM) (Pal ve Mather 2005; Mountrakis ve ark., 2011; Pal ve Foody 2012; Hawryło, 2018; Chakhar ve ark., 2021), Karar Ağacı (Pal ve Mather 2003, Trier ve ark., 2018; Zhang ve ark., 2020) ve Arttırılmış Karar Ağacı (Freund and Schapire 1997; Chan ve Paelinckx 2008) makine öğrenme algoritmaları bitki türünün tespitinde başarılı bir biçimde kullanılmıştır.

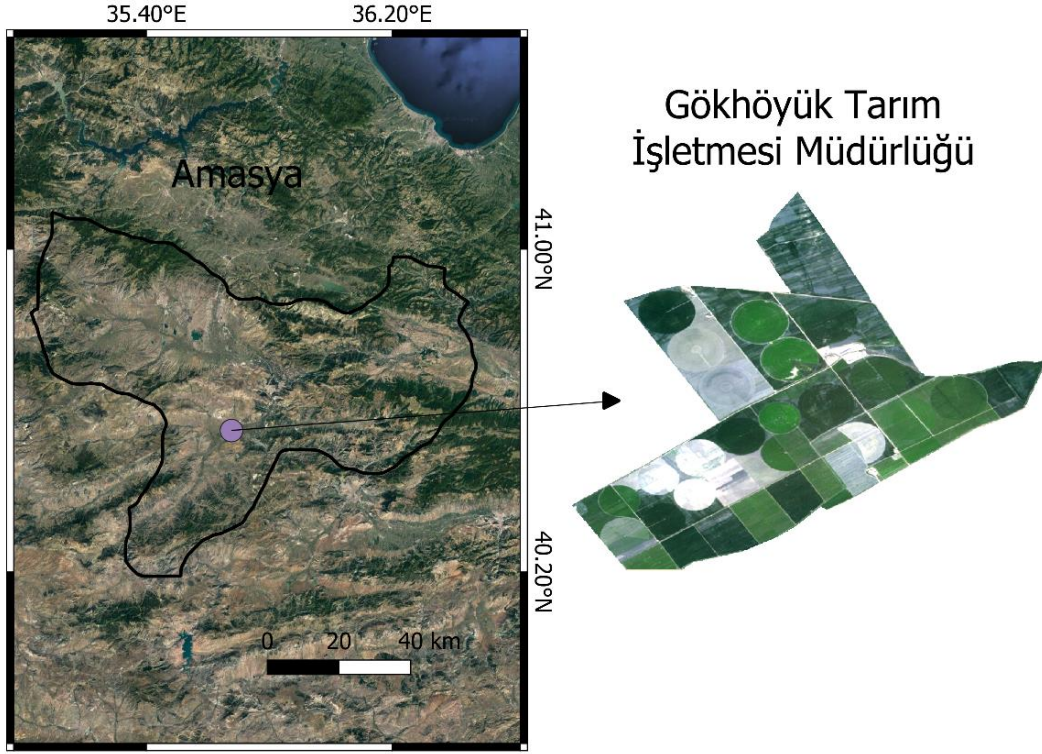
Literatürde MLC yöntemi ile bitki türü sınıflandırması üzerine oldukça fazla çalışma olmasına rağmen, makine öğrenmesi ile bitki türü sınıflandırması üzerine yapılmış araştırma oldukça azdır. Bu nedenle bu çalışmanın amacı parametrik olmayan RO, DVM ve K-NN makine öğrenme algoritmalarının bitki türü sınıflandırmasındaki başarısının ortaya konulmasıdır. Bu amaçla Amasya-Gökhöyük Tarım İşletmesi’nde 2020 yılı bitkisel üretim sezonunda yetiştirilen bitki türleri Sentinel 2 uydu görüntüleri ve makine öğrenme algoritmaları kullanılarak sınıflandırılmış ve elde edilen sonuçların doğruluğu değerlendirilmiştir.

## **Materyal ve Yöntem**

### **Çalışma Alanı**

Çalışma 2020 yılında Amasya ilinde yer alan Gökhöyük Tarım İşletmesi Müdürlüğü tarımsal üretim arazilerinde gerçekleştirilmiştir. Çalışma alanı Thornthwaite iklim sınıflandırmasına göre yarı kurak iklim özelliği göstermektedir (MGM, 2020). Uzun yıllar iklim verilerine göre ortalama yıllık toplam yağış miktarı 463.7 mm, ortalama en yüksek sıcaklık ve en düşük sıcaklıklar sırasıyla Ağustos ve Ocak aylarında 31.4°C ve -0.8 °C olarak gerçekleşmiştir (MGM, 2020).

Araştırma alanındaki bitkilerin sulanmasında dairesel hareketli ve klasik yağmurlama sulama sistemleri kullanılmaktadır. Sulama suyu işletme içerisindeki mevcut derin kuyulardan ve işletmenin yakınından geçen bir sulama kanalından temin edilmektedir. İşletmede başlıca yetiştirilen tarım ürünleri; buğday, fiğ, kenevir, nohut, yonca ve silajlık mısırdır.



Şekil 1. Çalışma alanının konumu ve genel görünümü

#### Çalışmada Kullanılan Uydu Görüntüleri

Çalışmada 2020 yılına ait 7 adet (18 Nisan, 18 Mayıs, 17 Haziran, 7 Temmuz, 11 Ağustos, 15 Eylül) Sentinel-2 uydu görüntüsü kullanılmıştır. Sentinel uydu sistemi Avrupa Uzay Ajansı (ESA) tarafından Copernicus programının ihtiyaçları doğrultusunda hazırlanmıştır. Sentinel-2 uydusu elektromanyetik spektrumun görünür ve infrared bölgelerinde toplam 13 adet multispektral bantta görüntü kaydetmektedir. Uydu görüntüleri Level-1C ve Level-2A seviyelerinde kullanıcılara sunulmaktadır. Level-1C, atmosfer üzeri (Top-of-Atmosphere) spektral yansımaya değerini sunmaktadır. Level-2A ise atmosfer altındaki (Bottom-of-Atmosphere), geometrik ve spektral doğrulamaları yapılmış yansımaya değerini sunmaktadır. Çalışmada bitki türü sınıflandırılmasında Level-2A seviyesindeki uydu görüntüleri kullanılmıştır. Level-1C seviyesindeki görüntüler ise öncelikle Level-2A seviyesine dönüştürülüp daha sonra sınıflandırmada kullanılmıştır.

Makine öğrenme eğitiminden önce, Sentinel 2 uydusuna ait hangi bantların modelde kullanılacağı belirlenmiştir. Söz konusu işlem her bir uydu verisi ile bitki türünü temsil eden kod arasındaki korelasyon analizi ile gerçekleştirilmiştir. Bu seçim işlemi sonucunda makine öğrenme algoritmalarının eğitim ve test aşamalarında Sentinel 2 uydusuna ait Bant 3, Bant 5, Bant 6, Bant 7, Bant 8, Bant 8A ve Normalize Edilmiş Vejetatif Değişim İndeksi (NDVI) kullanılmıştır. NDVI, tarımsal uzaktan algılama uygulamalarında en sık kullanılan ve bitki gelişim düzeyinin en başarılı bir biçimde gösterildiği vejetasyon indekslerinden birisidir. Bu nedenle bu çalışmada Sentinel-2 uydu görüntülerinin bazı bantlarına ek olarak NDVI değerleride veri setine dahil edilmiştir. NDVI, elektromanyetik spektrumun yakın kızıl ötesi ve kırmızı bölgesindeki yansımaya değerlerinin normalize edilmesiyle hesaplanmaktadır. Sentinel-2 uydu görüntülerinde NDVI hesabı Eşitlik 1' e göre gerçekleştirilmiştir.

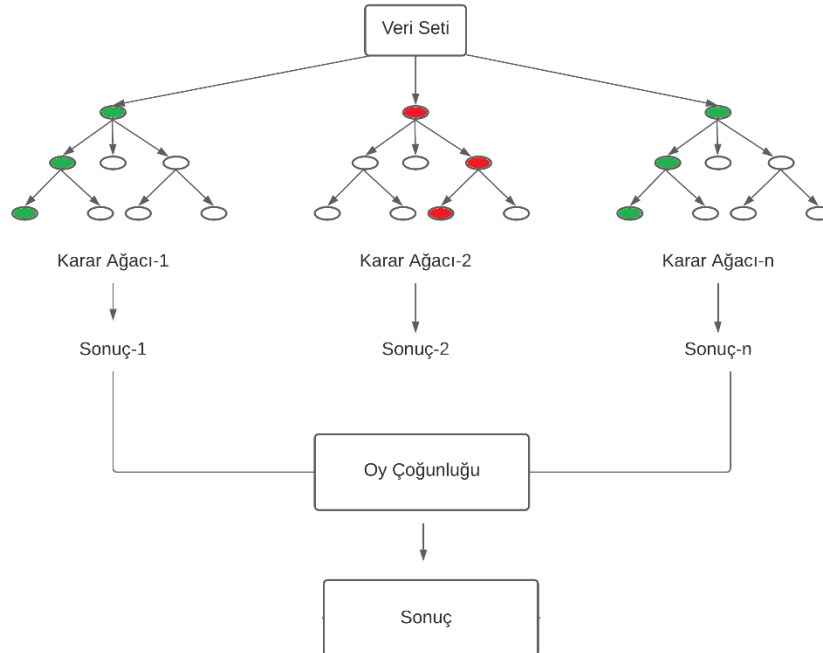
$$NDVI = \frac{(Bant\ 8 - Bant\ 4)}{(Bant\ 8 + Bant\ 4)} \quad (\text{Eşitlik 1})$$

### Makine Öğrenmesi Algoritmaları

Son yıllarda makine öğrenmesi ile sınıflandırma işlemi uzaktan algılama çalışmalarının odak noktası haline gelmiştir (Pal ve Mather, 2003). Makine öğrenmesi ile sınıflandırma işleminin ilk adımında hangi algoritmanın seçileceği sorusu ortaya çıkmaktadır. Literatürde çok sayıda makine öğrenme algoritmalarının olması ve yapılan sınıflandırma araştırmalarında henüz geliştirilebilir bir sonuç elde edilememiş olması, söz konusu algoritmaların seçimini zorlaştırmaktadır. Bu çalışmada literatürde sıklıkla tercih edilen RO, DVM ve K-NN makine öğrenmesi algoritmaları bitki türlerinin sınıflandırılması amacıyla kullanılmıştır.

#### Rassal Orman (RO)

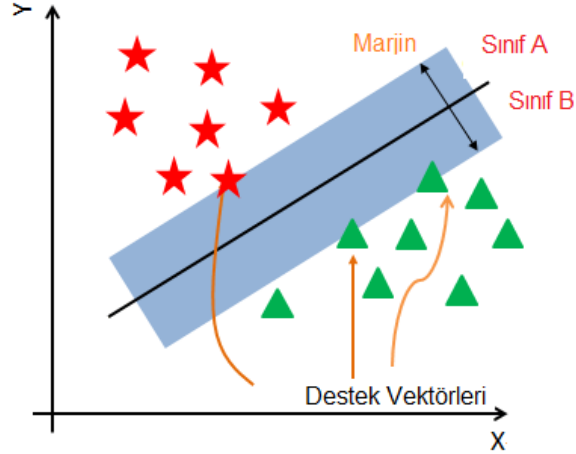
RO, veriyi birden fazla alt setlere bölerek karar ağaçları oluşturan bir makine öğrenme algoritmasıdır. Bu yöntem ile her bir karar ağacından ayrı tahminler elde edilir (Breiman, 2001). Oluşturulan bu karar ağaçlarındaki sonuçlarının çoğunluğu dikkate alınarak sınıflandırma işlemi gerçekleştirilmektedir (Şekil 2). RO algoritmasında birbiri ile aynı olmayan veri setleri kullanılarak makine eğitimi gerçekleştirildiği için aşırı öğrenme problemi oldukça düşüktür (Watts and Lawrence, 2008). RO ile bitki türlerinin sınıflandırmasında detayları Scikit-learn (2020)' de verilen yöntemler kullanılmıştır.



Şekil 2. Rassal Orman Algoritmasının Genel Gösterim Şeması

#### Destek Vektör Makinesi Sınıflandırması (DVM)

DVM, temelde farklı sınıflara ait verilerin doğrusal bir biçimde birbirinden ayrılmasında kullanılan bir makine öğrenme algoritmasıdır (Kumar ve ark., 2015). DVM’de ilk olarak verinin geometrik özellikleri belirlenir ve iki sınıfın birbirinden doğrusal olarak ayrışmasını sağlayan sınır belirlenir. Daha sonra bu sınırın iki tarafında bulunan ve marjin adı verilen karar sınırları belirlenir (Şekil 3). DVM’ de amaç karar sınırlarının en yüksek düzeye çıkartılmasıdır (Vapnik 1998; Huang ve ark., 2002). DVM ile bitki türlerinin sınıflandırma detayları Scikit-learn (2020)' de verilen yöntemler kullanılmıştır.



Şekil 3. Destek Vektör Makinesi Algoritmasının Genel Gösterim Şeması

### K-En Yakın Komşu Sınıflandırması (K-NN)

K-NN parametrik olmayan bir makine öğrenme algoritmasıdır. Bu yöntemde eğitim aşaması bulunmamaktadır. K-NN sınıflandırmasında, eğitim verilerini öğrenmek yerine, tüm veri seti algoritma tarafından ezberlenir. Bu yöntem ile sınıflandırma yapılmak istenildiğinde veri setine eklenecek olan yeni verinin, mevcut verilere göre uzaklığı hesaplanıp en yakın K adet komşu belirlenir. Sonraki adımda ise komşu elemanların uzaklık mesafeleri belirlenmektedir. Komşu verilere ait uzaklıklar sıralanır ve yeni gelen değere en uygun olan sınıf seçilmektedir (Maxwell ve ark., 2018). K-NN ile bitki türlerinin sınıflandırmasında detayları Scikit-learn (2020)' de verilen yöntemler kullanılmıştır.

### Çalışmada Kullanılan Yazılımlar

Sentinel 2 Level-1C görüntülerinden Level-2A görüntülerinin elde edilmesinde Sen2Cor yazılımı kullanılmıştır. Raster formatındaki uydu görüntülerinin işlenmesi, görselleştirilmesi ve analiz edilmesinde Quantum Geographic Information System (QGIS) 3.16.1 yazılımı kullanılmıştır. Ayrıca QGIS yazılımı ile her bir bitki türü için yer verilerine göre vektörler oluşturulmuştur. Bu vektörler kullanılarak görüntüler kesilmiş ve “.XYZ” formatında kaydedilmiştir. Makine öğrenmesi algoritmalarının eğitilmesi, test edilmesi ve veriler arasında öznitelik seçimi işlemlerinin gerçekleştirilmesinde Scikit-learn kütüphanesi (Pedregosa ve ark., 2011) kullanılmıştır.

### Veri Setinin Hazırlanması

Sentinel 2 uydu görüntülerinden elde edilen spektral veriler her bir bitki sınıfı için ayrı olarak etiketlenmiş ve her bir tarihe ait veriler alt alta birleştirilerek veri seti oluşturulmuştur. Bu çalışmada toplam 712.026 adet görüntü hücresinden elde edilen verilerin %70' i eğitim (496.701 adet) ve %30'u test (215.325 adet) verisi olarak kullanılmıştır. Veri setinin eğitim ve test olarak ayrıştırılması işleminde, veriler Scikit-learn kütüphanesi tarafından otomatik olarak seçilmiştir. Bitki sınıflarına göre seçilen eğitim ve test verileri sayıları Çizelge 1' de verilmiştir.

Çizelge 1. Çalışmada kullanılan bitki türü sınıfları ve bu sınıflara ait eğitim, test ve toplam veri sayıları

Bitki Türü	Eğitim Veri Sayısı	Test Veri Sayısı	Toplam Veri
Buğday	131.888	64.636	196.524
Fiğ	107.424	52.602	160.026
Kenevir	39.673	19.607	59.28
Yeni Tesis Yonca	46.175	23.017	69.192
Yonca	75.847	37.601	113.448
Mera	66.140	3.184	69.324
Ş.Pancarı	12.105	6.069	18.174
Soğan	17.449	8.609	26.058
<b>Toplam</b>	<b>496.701</b>	<b>215.325</b>	<b>712.026</b>



### İstatistiksel Değerlendirme

Makine öğrenmesi algoritmaları ile sınıflandırma sonuçlarının değerlendirilmesinde hata matrisleri, doğruluk (accuracy) ve F1-Skor değerleri kullanılmıştır. Hata matrisi iki veya daha fazla sınıfın olduğu bir veri setinde makine öğrenme algoritmalarının performansını ölçmede kullanılan bir tablodur (Çizelge 2). Bu tablo gerçek değerlere karşın tahmin edilen değerlerin olduğu bir yapıdır.

Çizelge 2. Genel Karmaşıklık Matrisi Tablosu

Tahmin Edilen Değerler		Gerçek Değerler	
		Pozitif (1)	Negatif (0)
Pozitif (1)	Pozitif (1)	TP	FP
	Negatif (0)	FN	TN

Çizelgede bulunan; TN: Negatif Doğru, TP: Pozitif Doğru, FP: Pozitif yanlış, FN: Negatif yanlış anlamına gelmektedir.

Doğruluk ve F1- Skor değeri Eşitlik 2 ve 3 'e göre hesaplanmıştır.

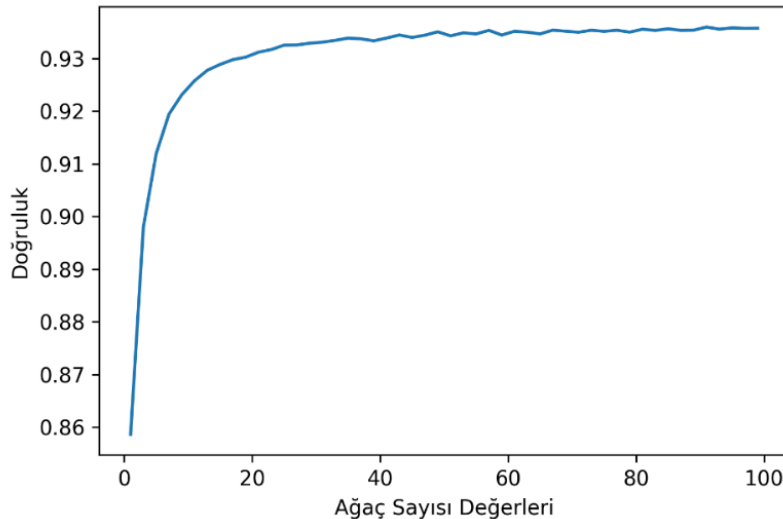
$$\text{Doğruluk} = \frac{(TP+TN)}{(\text{Toplam Veri Sayısı})} \quad (\text{Eşitlik 2})$$

$$F1 - \text{Skoru} = \frac{2 \times TP}{2 \times (TP+FP+FN)} \quad (\text{Eşitlik 3})$$

### Bulgular ve Tartışma

#### Rassal Orman Sınıflandırması

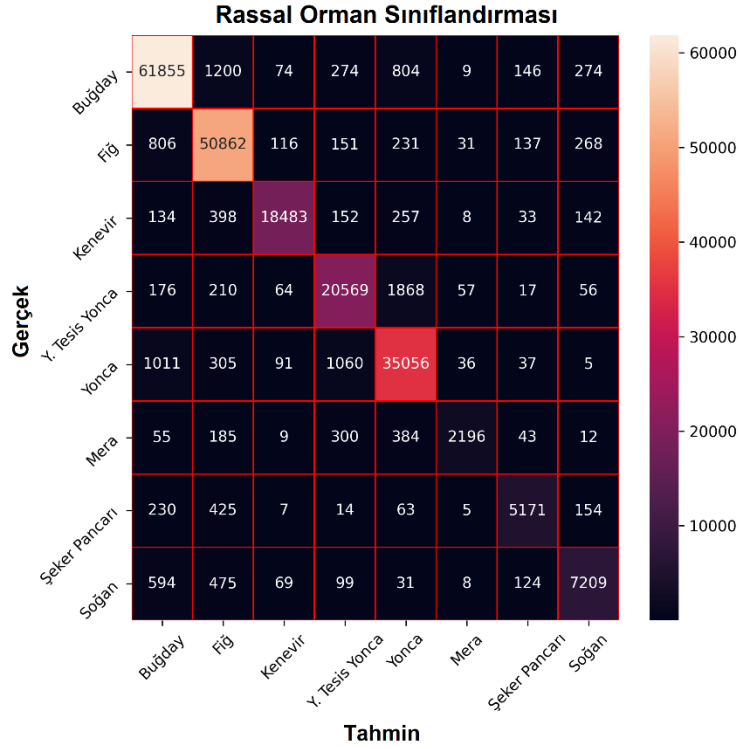
RO makine öğrenme algoritmasının başarısı kullanıcı tarafından belirlenen ağaç sayısına ( $n_{estimators}$ ) bağlı olarak değişim göstermektedir (Breiman, 2001). Bu çalışmada ilk olarak RO algoritması eğitim verileri üzerinde bir döngü içerisinde çalıştırılıp, en uygun ağaç sayısı belirlenmiştir. Döngü sonucunda elde edilen doğruluk (accuracy) değerleri Şekil 4 'te verilmiştir. Şekil 4 incelendiğinde en yüksek doğruluk, ağaç sayısının 96 (doğruluk değeri= 0.9358) olarak kullanıldığı döngüde elde edilmiştir. Buna karşın en düşük doğruluk oranı ise 0.8597 (ağaç sayısı= 1) olarak hesaplanmıştır.



Şekil 4. Rassal Orman Algoritmasında Ağaç Sayısı Değişiminin Doğruluk Seviyesine Etkisi

En uygun karar ağaç sayısının kullanıldığı RO modelinin etkinliğinin belirlenmesi amacıyla hata matrisi hesaplanmış ve sonuçlar Şekil 5'te verilmiştir. Hata matrisi incelendiğinde, RO algoritmasının sınıflandırma başarısı mera dışındaki tüm bitki örtülerinde %85'in üzerinde

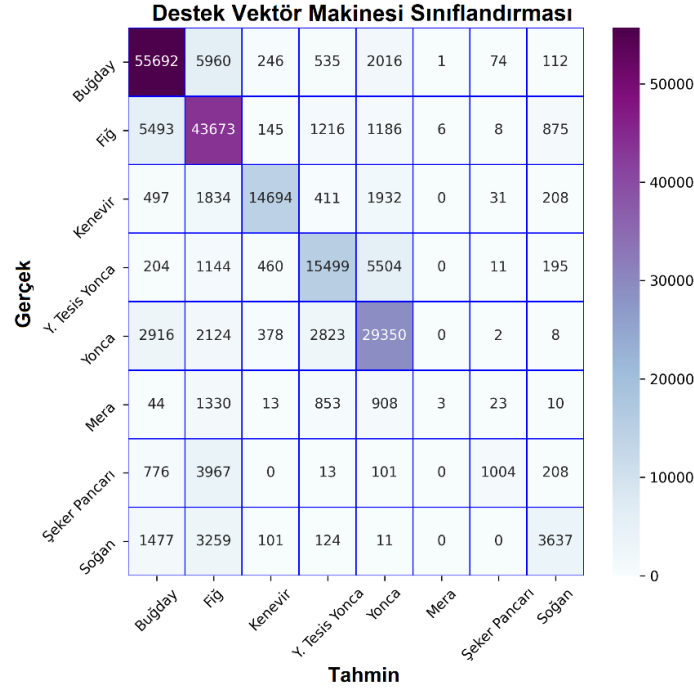
gerçekleşmiştir. En başarılı sınıflandırma fiğ bitkisinde gerçekleşirken, en düşük başarı ise meraların sınıflandırılmasında gerçekleşmiştir. Fiğ bitkilerine ait toplam 52.602 adet test verisinin 50.862 adeti doğru tahmin edilirken, 1.740 adet veri hatalı tahmin edilmiştir. Fiğ bitkisindeki hatalı gerçekleştirilen sınıflandırma en fazla buğday sınıfında (%46) gerçekleşirken, en düşük hata mera sınıfında (%1) meydana gelmiştir. Meralara ait 3.184 adet test verisinin 2.196 tanesi doğru tahmin edilirken, 988 tanesi ise hatalı tahmin edilmiştir.



Şekil 5. Rassal Orman Sınıflandırmasına ait Hata Matrisi Sonuçları

#### Destek Vektör Makinesi Sınıflandırması

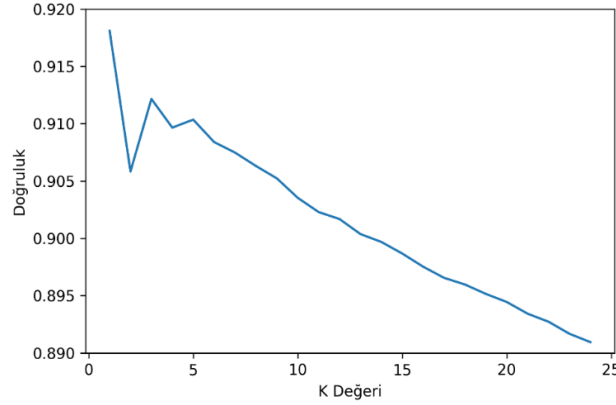
DVM makine öğrenme algoritması ile çalışma alanına ait bitki sınıflandırılması gerçekleştirilmiş ve DVM algoritmasının bitki sınıflandırmadaki başarısını gösteren hata matrisi Şekil 6'da verilmiştir. En başarılı sınıflandırma %86 başarı yüzdesi ile buğdayda gerçekleşirken, en düşük başarı oranı ise meraların sınıflandırılmasında gerçekleştirilmiştir. Mera sınıfına ait 3.184 adet verinin sadece 3 tanesi doğru sınıflandırılmıştır. Benzer şekilde şekerpancarı (%17) ve soğan (%42) sınıfları için gerçekleştirilen sınıflandırma tahmini oldukça düşük başarı ile gerçekleşmiştir. Bu sınıflardaki tahmin başarısının düşük olmasının en önemli nedeni eğitim setindeki örnek sayılarının az olmasıdır. Pal and Foody (2010) tarafından yapılan bir çalışmada, SVM algoritması, veri setinin boyutuna ve eğitim setindeki örnek sayısına oldukça hassas olduğu belirtilmiştir. Kumar et al. (2015)' de SVM algoritmasının başarısı eğitim setindeki veri sayısı arttıkça düzenli olarak artış gösterdiği bildirilmiştir.



Şekil 6. Destek Vektör Makinesi Sınıflandırmasına ait Hata Matrisi Sonuçları

#### K-En Yakın Komşu (K-NN) Sınıflandırması

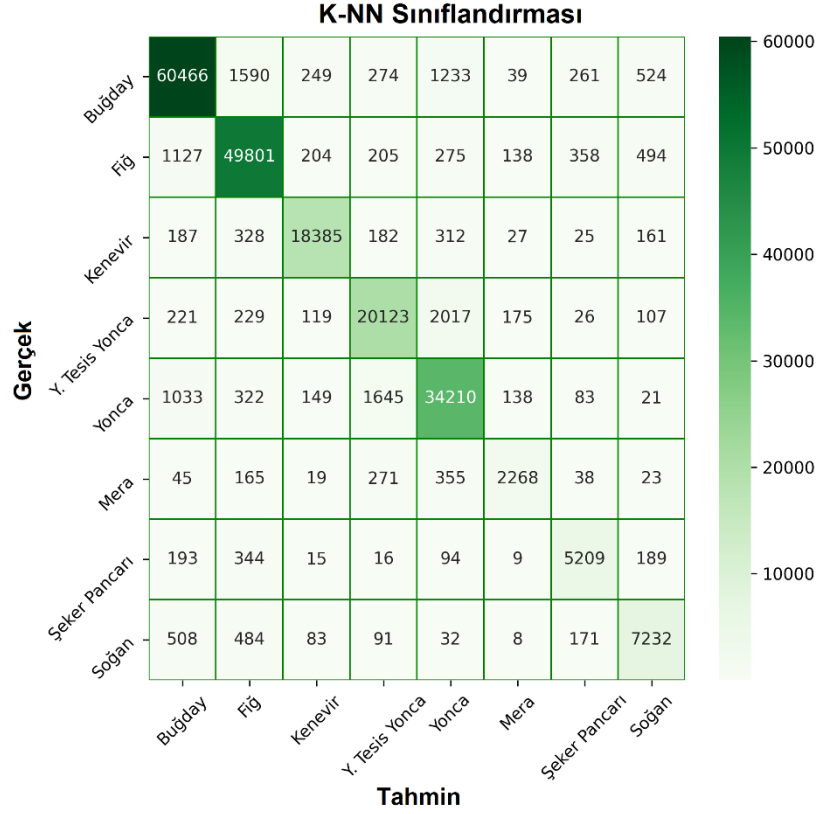
K-NN makine öğrenme algoritmasında tahmin edilecek verinin, mevcut olan verilere göre uzaklığı belirlenip, K sayıda yakın komşuluğu dikkate alınır. Bu nedenle optimum K sayısının belirlenmesi için K-NN algoritması eğitim verilerinde bir döngü içerisinde çalıştırılmış ve sonuçlar Şekil 7’ de verilmiştir. Buna göre K değeri arttıkça K-NN algoritmasının doğruluk değerinin azaldığı belirlenmiştir.



Şekil 7. K-NN Algoritmasında “K” Değerinin Doğruluk Seviyesine Etkisi

En uygun K değeri (1) kullanılarak K-NN makine öğrenmesi algoritması test verileri üzerinde çalıştırılıp hata matrisi hesaplanmış ve sonuçlar Şekil 8’de verilmiştir. Şekil 8 incelendiğinde toplam 215.325 adet test verisinin 197.694 tanesi doğru olarak tahmin edilirken, 17.631 adet veri hatalı tahmin edilmiştir. RO algoritmasında olduğu gibi K-NN algoritmasında da en başarılı sınıflandırma fiğ bitkisi sınıfında gerçekleştirilirken, en düşük başarı meralarda elde edilmiştir. Fiğ bitki sınıfında toplam 49.801 adet veri doğru tahmin edilirken, mera sınıfında sadece 2.268 adet veri doğru tahmin edilmiştir. K-NN makine öğrenme algoritması ile buğday, fiğ ve kenevir sınıfları sırasıyla %94, %95 ve %94 başarı ile tahmin edilirken mera, şeker pancarı ve soğan sınıfları sırasıyla %71, %86 ve %84 başarı ile tahmin edilmiştir.

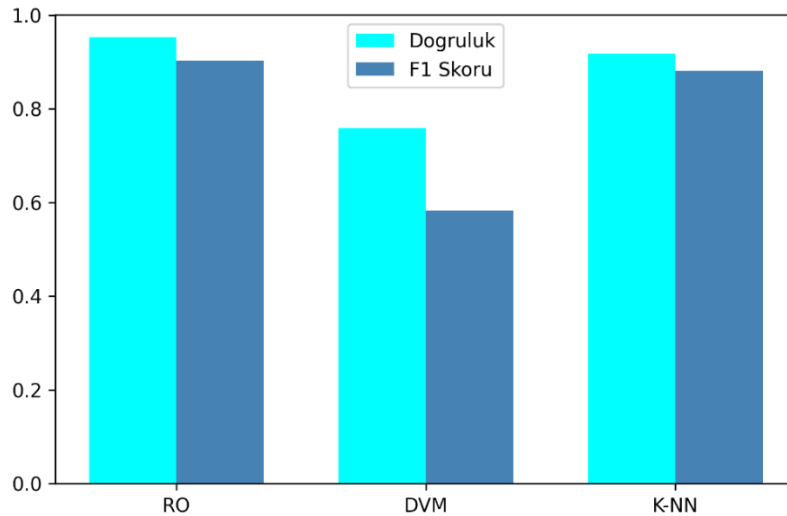




Şekil 8. K-NN Sınıflandırmasına ait Hata Matrisi Sonuçları

### Makine Öğrenmesi Algoritmalarının Karşılaştırılması

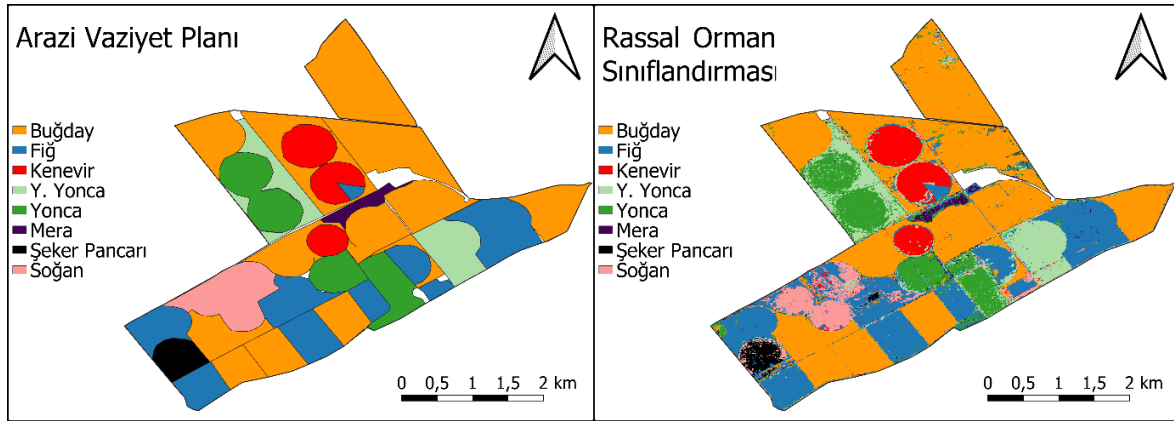
Çalışmada kullanılan makine öğrenme algoritmalarının genel başarı düzeylerinin ortaya konulması için doğruluk ve F1 skor değerleri hesaplanmış ve sonuçlar Şekil 9’da verilmiştir. Şekil incelendiğinde en yüksek doğruluk (0.953) ve F1 skor (0.903) değerleri RO algoritmasından elde edilmiştir. En düşük başarı ise DVM algoritması ile gerçekleştirilen sınıflandırma sonucunda elde edilmiştir. K-NN algoritması ile RO algoritmasına yakın sonuçlar elde edilmiş ve doğruluk ve F1-skor değeri sırasıyla 0.918 ve 0.881 olarak hesaplanmıştır.



Şekil 9. Çalışmada Kullanılan Makine Öğrenmesi Algoritmalarının Doğruluk ve F1 Skoru Karşılaştırma Grafiği

Sınıflandırma sonucu en yüksek başarı düzeyine sahip RO makine öğrenme modeli kullanılarak oluşturulan bitki türü sınıflandırma haritası Şekil 10’da verilmiştir. Harita görsel olarak incelendiğinde genel olarak tüm bitkilerin başarılı bir biçimde ayrıştığı görülmektedir. Ayrıca ülkemiz için stratejik bitkilerden biri haline gelen kenevir bitkisi, RO makine öğrenme algoritması ile oldukça başarılı bir biçimde sınıflandırılmıştır. Siyah renk ile temsil edilen şeker pancarı parseli genel olarak başarılı bir biçimde sınıflandırılmasına rağmen, yetiştirme alanı içerisindeki bazı görüntü hücreleri soğan olarak hatalı sınıflandırılmıştır. Benzer şekilde mera alanlarının bir kısmı yonca sınıfı olarak tahmin edilmiştir. Bu çalışmada makine öğrenme algoritmaları kullanılarak bitki türü sınıflandırma işlemi genel olarak başarılı bir biçimde gerçekleştirilmiştir. Söz konusu teknik uydu görüntülerine uygulanarak geniş alanlarda bitkilerin izlenmesinde kullanılabilir. Ayrıca izin alınmadan kenevir yetiştiriciliği yapılan tarım alanları makine öğrenmesi algoritmaları ile başarılı bir biçimde tespit edilebilir.

Makine öğrenme algoritmalarının işlemi gerçekleştirme süreleri dikkate alındığında RO algoritması ile oluşturulan model test verileri üzerinde 123.5 sn’de sonuç üretilirken, bu süre K-NN algoritması ile sadece 23.6 sn’de tamamlanmıştır. Doğruluk ve F1-skor değerlerinin birbirine yakın gerçekleşmesi göz önünde bulundurulduğunda, örnek sayısının oldukça fazla olduğu çalışmalarda K-NN algoritması kullanılarak hem daha hızlı hem de daha ekonomik olarak model sonuçları elde edilebilir.



Şekil 10. Arazi Vaziyet Planı ve Rassal Orman Sınıflandırma Haritası

### Sonuç ve Öneriler

Bitki türünün uzaktan algılama teknikleri ile belirlenmesi, tarımsal politikaların oluşturulması ve bu politikaların izlenmesi bakımından oldukça kıymetlidir. Bu çalışmada Gökhöyük Tarım İşletmesi Müdürlüğü tarımsal üretim alanlarında yetiştirilen bitki türleri Sentine-2 uydu görüntüsü ve makine öğrenme algoritmalarıyla tahmin edilmiş ve elde edilen sonuçlar yer doğrularıyla değerlendirilmiştir. Çalışmada literatürde sıklıkla tercih edilen makine öğrenme algoritmalarından RO, DVM ve K-NN kullanılmıştır. Araştırma sonuçlarına göre bitki türleri RO, DVM ve K-NN makine öğrenmesi algoritmalarıyla sırasıyla %95.3, %75.9 ve %91.8 doğruluk değeri ile tahmin edilmiştir. Elde edilen bu sonuçlar, bitki türlerinin makine öğrenme algoritmalarıyla başarılı bir biçimde sınıflandırılabilirliğini göstermektedir. Bu yöntem ile bitkisel üretim alanlarının takibi ve kontrolü başarılı bir biçimde gerçekleştirilebilir. Bölgesel veya ülkesel boyutta yapılacak bitki türü sınıflandırması ile bitkisel üretim alanları etkin bir biçimde belirlenip, tarımsal politikaların geliştirilmesine katkıda bulunabilir. Ayrıca uzaktan algılanmış görüntüler ile makine öğrenmesi algoritmaları kullanılarak izinsiz yetiştirilen kenevir alanları belirlenebilir. Gelecekte farklı uydu görüntüleri ve farklı makine öğrenme algoritmalarıyla geniş alanları kapsayan çalışmaların yapılmasına ihtiyaç bulunmaktadır. Ayrıca makine öğrenmesi algoritmalarına ek olarak derin öğrenme ve yapay sinir ağlarının bitki sınıflandırmasındaki başarısının belirlenmesi üzerine araştırmaların yapılmasına ihtiyaç bulunmaktadır.



### Kaynaklar

- Baker, C. 1987. Changes in financial markets and their effects on agriculture. Federal Reserve Bank of St. Louis Review.
- Breiman L. 2001. Random forests, machine learning. Kluwer Academic Publishers. 45 (1): 5-32.
- Chakhar, A., Hernández-López, D., Ballesteros, R., Moreno, M. A. 2021. Improving the accuracy of multiple algorithms for crop classification by integrating sentinel-1 observations with sentinel-2 data. Remote Sensing. 13(2), 243.
- Chan, J. C., Paelinckx, D. 2008. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment. 112 (6): 2999-3011.
- Doğan, Z., Arslan, S., Berkman, A. N. 2015. Türkiye'de tarım sektörünün iktisadi gelişimi ve sorunları: tarihsel bir bakış. Academic Review of Economics & Administrative Sciences. 8(1): 1308-4208.
- Freund, Y., R. E. Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences. 55: 119–139.
- Hawryło, P., Bednarz, B., Wężyk, P., Szostak, M., 2018. Estimating defoliation of scots pine stands using machine learning methods and vegetation indices of sentinel-2. Eur. J. Remote Sens. 51: 194–204.
- Homer, C, Huang, C., Yang, L., Wylie, B., Coan, M. 2004. Development of a 2001 national land-cover database for the United States. Photogrammetric Engineering. 70(7): 829.
- Hooda, R. S., Yadav, M., Kalubarme, M. H. 2006. Wheat production estimation using remote sensing data: An Indian experience. Workshop Proceedings: Remote Sensing Support to Crop Yield Forecast and Area Estimates, Stresa, Italy. 30 Nov.–1 Dec. 2006.
- Huang, C., L. S. Davis, J. R. G. Townshend. 2002. An Assessment of support vector machines for land cover classification. International Journal of Remote Sensing. 23: 725– 749.
- Jay, S., Lawrence, R., Repasky, K., Keith, C. 2009. Invasive species mapping using low cost hyperspectral imagery. ASPRS 2009 Annual Conference. Baltimore, Maryland.
- Kumar, P., Gupta, D. K., Mishra, V. N., Prasad, R. 2015. Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using LISS IV data. International Journal of Remote Sensing. 36(6): 1604-1617.
- Li, L., Zheng, X., Zhao, K., Li, X., Meng, Z., Su, C. 2020. Potential evaluation of high spatial resolution multi-spectral images based on unmanned aerial vehicle in accurate recognition of crop types. Journal of the Indian Society of Remote Sensing.
- Maxwell, A. E., Warner, T. A., Fang, F. 2018. Implementation of machine-learning classification in remote sensing: An applied review. IJRS. 39(9): 2784-2817
- MGM, 2020. Meteoroloji Genel Müdürlüğü, Mevbis Sistemi.
- Mountrakis, G., Im, J., Ogole, C. 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing. 66(3): 247-259.
- Ok, A. O., Akar, Ö., Güngör, O. 2012. Evaluation of random forest method for agricultural crop classification. European Journal of Remote Sensing. 45(1): 421-432.
- Pal, M., Foody, G. M. 2010. Feature selection for classification of hyperspectral data by SVM. IEEE Transactions on Geoscience and Remote Sensing. 48(5): 2297-2307.
- Pal, M., Foody, G. M. 2012. Evaluation of SVM, RVM and SMLR for accurate image classification with limited ground data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 5(5): 1344-1355.
- 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): 554-565.
- Pal, M., Mather, P. M. 2015. Support vector machines for classification in remote sensing. International journal of remote sensing. 26(5): 1007-1011.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. 2011. Scikit-learn: machine learning in Python. Journal of machine Learning research. 12: 2825-2830.
- Prasad, A. M., Iverson, L. R., Liaw, A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems. 9: 181-199.
- Scikit-learn. 2020. Scikit-learn: Machine learning in Python. <https://scikit-learn.org>
- Trier, Ø. D., Salberg, A. B., Kermit, M., Rudjord, Ø., Gobakken, T., Næsset, E., Aarsten, D. 2018. Tree species classification in Norway from airborne hyperspectral and airborne laser scanning data. European Journal of Remote Sensing. 51(1), 336-351.



- Uzundumlu, A. S. 2012. Tarım sektörünün ülke ekonomisindeki yeri ve önemi. *Alinteri Journal of Agriculture Sciences*. 22(1):1.
- Vapnik, V. N. 1998. *Statistical learning theory*. New York: Wiley.
- Waske, B., Braun, M. 2009. Classifier ensembles for land cover mapping using multi-temporal SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*. 64: 450–457.
- Watts, J. D., Lawrence, R. L. 2008. Merging random forest classification with an object-oriented approach for analysis of agricultural lands. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXXVII(B7):1.
- Yang, C., Everitt, J. H., Murden, D. 2011. Evaluating high resolution SPOT 5 satellite imagery for crop identification. *Computers and Electronics in Agriculture*. 75 (2): 347.
- Zhang, H., Kang, J., Xu, X., Zhang, L. 2020. Accessing the temporal and spectral features in crop type mapping using multi-temporal Sentinel-2 imagery: A case study of Yi'an County, Heilongjiang province, China. *Computers and Electronics in Agriculture*. 176(05618).