
Araştırma Makalesi / Research Article

Makine Öğrenmesi Algoritmaları Kullanılarak Kayısı İç Çekirdeklerinin Sınıflandırılması

Fatih Ahmet ŞENEL*

*Süleyman Demirel Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Isparta
(ORCID: 0000-0003-1918-7277)*

Öz

Kayısı çekirdeği üretimi ve tüketimi fazla olan bir gıda ürünüdür. Kayısı iç çekirdeklerinin makine öğrenmesi algoritmaları kullanılarak, tatlı veya acı olarak sınıflandırılması bu çalışmanın konusunu oluşturmaktadır. Hem tatlı hem de acı kayısı iç çekirdeği için talep miktarı oldukça fazladır. Depolama şartları gibi nedenlerden dolayı kayısı iç çekirdekleri zaman zaman birbirine karışabilmektedir. Bu durum tüketiciler tarafından istenmeyen bir durumdur. Kayısı iç çekirdeğinin ayrıştırılması, gözle her zaman mümkün olmamaktadır. Bu çalışmanın amacı, insan faktörünü ortadan kaldıracak bir sınıflandırma yönteminin geliştirilmesidir. Bu sınıflandırma işlemi için k En Yakın Komşu, Destek Vektör Makinesi, Karar Ağacı, Rasgele Orman, Adaptive Boosting, Gaussian Naive Bayes ve Çok Katmanlı Algılayıcı algoritmaları kullanılmıştır. Yeterli sayıda öznelik ile algoritmaların yarısından fazlası sınıflandırma işlemi %100 başarı ile elde edebilmektedir. En az sayıda öznelik kullanarak en iyi başarı Rasgele Orman algoritması ile elde edilmiştir. Sonuçlar göstermiştir ki, kayısı iç çekirdeklerinin sınıflandırılması işlemi makine öğrenmesi algoritmaları ile başarılı bir şekilde gerçekleştirilebilmektedir.

Anahtar kelimeler: Kayısı iç çekirdeği, makine öğrenmesi, sınıflandırma.

Classification of Apricot Kernels by using Machine Learning Algorithms

Abstract

Apricot kernel is a food product with high production and consumption. The classification of apricot kernels as sweet or bitter using machine learning algorithms is the subject of this study. The demand for both sweet and bitter apricot kernels are quite high. Due to reasons such as storage conditions, apricot kernels can get mixed from time to time. This situation is undesirable for consumers. The separation of apricot kernels is not always possible with the eyes. This study aims to develop a classification method that can eliminate the human factor. For this classification process, k Nearest Neighbor, Support Vector Machine, Decision Tree, Random Forest, Adaptive Boosting, Gaussian Naive Bayes and Multilayer Perceptron algorithms have been used. More than half of the algorithms with a sufficient number of attributes can achieve the classification process with 100% success. The best performance has been obtained by the Random Forest algorithm using the minimum number of features. The results show that the classification of apricot kernels can be carried out successfully with machine learning algorithms.

Keywords: Apricot kernels, classification, machine learning.

1. Giriş

Türkiye kayısı üretiminde dünyada birinci sırada yer almaktadır. Ülkemizin kayısı üretiminin yaklaşık olarak %60'lık kısmı Malatya'da üretilmektedir [1]. Kayısı ticaretine bakıldığında, kayısının birçok farklı formda ticaretinin yapıldığı görülmektedir. Sofralık kayısı ticaretinin yanı sıra kurutulmuş kayısı ticareti de en önemli aşamalarındandır. Bununla birlikte, kayısının iç çekirdeği de bu ticaretin bir parçasını oluşturmaktadır. Kayısı iç çekirdeği; acı ve tatlı olmak üzere iki sınıfa ayrılmaktadır. Vitaminler bakımından zengin olması nedeniyle kayısı iç çekirdeği de günümüzde fazlaca

*Sorumlu yazar: fatihsenel@sdu.edu.tr

Geliş Tarihi: 03.07.2019, Kabul Tarihi: 20.04.2020

tüketilmektedir. Acı kayısı iç çekirdeği, kanser gibi bazı hastalıklara karşı tedavi amacıyla özellikle Çin gibi ülkelerde çokça tüketilmektedir [2]. Hem acı hem de tatlı kayısı iç çekirdekleri çoğunlukla farklı türde kayıslara ait olmasına karşın gerek depolanma ortamlarının aynı olması gerekse aynı ağaçlardan hem tatlı hem de acı kayısı iç çekirdeklerinin olması nedeniyle zaman zaman birbirlerine karışabilmektedir. Acı kayısı iç çekirdeğinin, hastalıklara iyi geldiği için tüketilmesinin yanı sıra, fazla tüketilmesinde de bazı sakıncalar ortaya çıkmaktadır [3,4]. Bu nedenle kayısı iç çekirdeğinin tüketilmesinde dikkatli davranmak gerekmektedir.

Yukarıda bahsedilen olumsuz durumlardan kaçınabilmek için, acı ve tatlı kayısı iç çekirdeklerinin birbirlerine karışmasının önüne geçmek gerekmektedir. Her ne kadar acı kayısı iç çekirdeği çoğu zaman daha açık renkli de olsa da birbirlerinden ayırt etmenin zor olacağı örnekler de bulunmaktadır. Ayrıca, insanlar tarafından bu ayırıştırma işleminin yapılması çok zahmetli olup, yorgunluğu bağlı dikkat bozulması gibi etkenlerden dolayı gözden kaçabilmektedir. Bu nedenle bu ayırıştırma işleminin bilgisayar ortamında görüntü işleme araçları ile tespitlerin yapılması ve ayırıştırılması en doğru yöntemdir. Günümüz bilgisayar teknolojisinin geldiği noktada artık görüntü işleme teknikleri ile çok daha hızlı ve yüksek doğrulukta sonuçlar elde edilebilmektedir.

Karhan ve arkadaşları [5], yaptıkları çalışmada görüntü işleme tekniklerini kullanarak kayısı üzerindeki yaprak delen hastalığının neden olduğu lekeleri tespit etmeyi amaçlamışlardır. Khojastehnazhand ve arkadaşları [6], kayısı meyvesini görüntü işleme tekniklerinden yararlanarak, olgunluk seviyesine göre üç farklı gruba sınıflandırma işlemini gerçekleştirmişlerdir. Her sınıflandırma işlemine ek olarak kayısların hacimlerinin tahmini işlemini de yapmışlardır. Hussain Hassan ve Nashat [7], görüntü işleme tekniklerinden yararlanarak zeytin meyvesi hastalık tespit ve sınıflandırma işlemini gerçekleştirmişlerdir. Alam ve arkadaşları [8], multispektral görüntüleri kullanarak kusurlu elmaları tanımlama işlemini gerçekleştirmişlerdir. Çalışmalarında %97'lik bir başarı oranı elde etmişlerdir. Ye ve arkadaşları [9], patateslerin üzerindeki çürük lekelerini hiperspektral görüntüleri kullanarak tespit etmeyi amaçlamışlardır. Çalışmalarında %100'lük bir başarı sonucu elde etmişlerdir. Wan ve arkadaşları [10], görüntü işleme tekniklerini kullanarak domateslerin olgunluk seviyelerini sınıflandırmışlardır. Sınıflandırma işlemi için geriye yayılım yapay sinir ağını kullanmışlardır. %99,31'lik bir başarı ile domatesleri doğru bir şekilde sınıflandırabilmişlerdir. Nasirahmadi ve Ashtiani [11], yaptıkları çalışmada kabuklu bademleri acı ve tatlı olarak sınıflandırma işlemini gerçekleştirmişlerdir. Badem resimlerinden yola çıkarak histogram özelliklerini kullanan Bag-of-Word yöntemi ile özneliklerini çıkarmışlardır. Çıkardıkları öznelikler ile üç farklı sınıflandırıcı kullanarak bademleri acı ve tatlı olarak sınıflandırmışlardır.

Thong ve arkadaşları [12], mango meyvesinin otomatik olarak sınıflandırıldığı ve yapay sinir ağları ile çalışan bir görüntü işleme sistemi geliştirmişlerdir. Mangoların renk, boyut, hasar görmüş gibi kriterlerini dikkate alarak sınıflandırma işlemini gerçekleştirmişlerdir. Görüntü üzerinden mangoların ağırlıklarını tahmin etmişlerdir. Luan ve arkadaşları [13], evrişimsel sinir ağlarını kullanarak ayçiçeği tohumunu sınıflandırmışlardır. Kullandıkları katmanlı mimari ile ağa girdi olarak ayçiçeği tohumlarının görüntülerini vererek, klasik sinir ağlarından daha iyi sonuçlar elde ettiklerini göstermişlerdir. Zeng ve arkadaşları armut meyvesinde bulunan yara ve ezikleri termal bir kameradan aldıkları görüntüler ile tespit etme işlemini gerçekleştirmişlerdir. Evrişimsel sinir ağlarına termal kameradan aldıkları görüntüleri girdi olarak vermişler ve yaraları, ezikleri tespit etmişlerdir. %99.25'lik bir doğrulukla başarılı sonuçlar elde etmişlerdir [14]. Zhao ve arkadaşları [15], nanegiller ailesinden olan sağlık ve gıda sektöründe kullanılan Perilla bitkisinin makine öğrenmesi ile sınıflandırılması işlemini gerçekleştirmişlerdir. Kamera ile aldıkları görüntülere bir matematiksel model geliştirerek modellemesini yapmışlardır. Elde ettikleri modeli kullanarak %93'lük bir doğru sınıflandırma sonucu elde etmişlerdir.

Bu çalışmada, makine öğrenmesi algoritmaları kullanılarak, kayısı iç çekirdeğinin sınıflandırılması işlemi gerçekleştirilmiştir. Kayısı iç çekirdekleri acı ve tatlı olarak iki farklı sınıfta yer almaktadır. Kayısı iç çekirdeklerinde renk ve boyut olarak farklılıklar bulunmakta olduğu için özneliklerde boyut ve renk içerecek şekilde oluşturulmuştur. Özneliklerin belirlenmesinden sonra öznelik seçme işlemi gerçekleştirilerek sınıflandırma işlemine etkisi az olan veya etkisi olmayan öznelikler çıkarılarak sınıflandırma işlemi yapılmıştır.

Bu çalışmanın ikinci bölümünde kullanılan yöntemler, öznelik oluşturma ve seçme işlemleri anlatılmıştır. Üçüncü bölümde araştırma bulgularına yer verilmiştir ve son bölümde ise elde edilen sonuçlar anlatılmıştır.

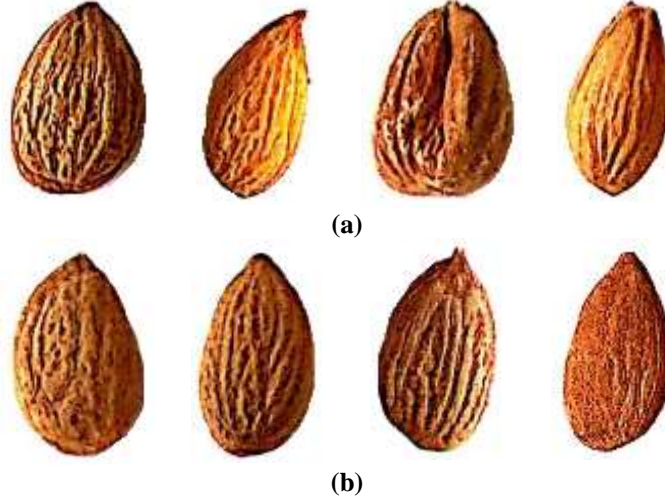
2. Materyal ve Metot

2.1. Veri Setinin Oluşturulması

Bu çalışmada kayısı iç çekirdeğinin sınıflandırılması işlemi gerçekleştirilmiştir. Sınıflandırma işleminin ilk aşaması olan veri setinin oluşturulurken her iki sınıfa ait (hem acı hem de tatlı) kayısı iç çekirdeği fotoğrafları çekilmiştir. Veri setinde uzmanlar tarafından sınıflandırılmış, 98 adet acı kayısı iç çekirdeği, 135 adet tatlı kayısı iç çekirdeği olmak üzere toplamda 237 adet kayısı iç çekirdeği verisi bulunmaktadır.

2.2. Özniteliklerin Oluşturulması

Öznitelik, verilerin sınıflara ayrıştırılması işlemi gerçekleştirebilecek, veriye ait özellikleri ifade etmektedir. Bu çalışmada kullanılan kayısı iç çekirdekleri incelendiğinde, acı ve tatlı olarak sınıflandırılan çekirdeklerin renk ve boyut olarak farklılıklar gösterdiği gözlemlenmiştir (Şekil 1). Bu nedenle öznitelikler oluşturulurken hem renk hem de boyut olarak özelliklerin belirlenmesi gerekmektedir. Boyut olarak incelendiğinde, fotoğrafların hem çekim açısı hem de büyüklük farklılıkları göz önünde bulundurulduğunda uzunluk olarak özelliklerin belirlenmesinin yanlışlıklara yol açacağı düşünülmüştür. Bu nedenle uzunluk olarak öznitelik belirlemek yerine oransal olarak özniteliklerin belirlenmesi daha doğru bir tercih olacaktır. Kayısı iç çekirdeğini çevreleyen ve alanı en az olan dikdörtgen bulunmuş, dikdörtgenin kısa kenarının uzun kenarına olan oranı öznitelik olarak belirlenmiştir. Boyut özneliğine ek olarak renk uzaylarına dönüşümler yapılarak toplamda 40 adet öznitelik kümesi oluşturulmuştur. Tablo 1’de çıkarılan öznitelikler verilmiştir.



Şekil 1. Örnek kayısı iç çekirdekleri görüntüsü (a) acı kayısı iç çekirdeği (b) tatlı kayısı iç çekirdeği

2.3. Öznitelik Seçme

Öznitelik seçme işlemi; çıkarılan öznitelik kümesinden, en iyi sonucu verecek bir öznitelik alt kümesinin oluşturulması aşamasıdır. Bu çalışmada özniteliklerin sonuca olan etkilerine göre skorlama işlemi yapılmış ve en güçlü öznitelikten en düşüğe doğru farklı sayılarda öznitelikler seçilmiştir. Bu çalışmada özniteliklerin skorlanması sonucunda en güçlüden en zayıfa doğru ilk 10 öznitelik Tablo 4’te verilmiştir. Öznitelik seçme işlemi yapılırken korelasyon tabanlı öznitelik seçme yöntemi uygulanmıştır. Uygulanan bu yöntemle, her bir öznitelik ile sınıf etiketi arasındaki korelasyonun maksimum olması ve seçilen özniteliklerin kendi aralarında ise minimum korelasyona sahip olmaları amaçlanmaktadır. Böylece seçilen öznitelik alt kümesi içinde sınıf etiketini tahmin ederken benzer özellikler gösteren özniteliklerden sadece bir tanesinin bulunması ve diğerlerinin elenmesi gerçekleştirilmiş olmaktadır. Seçilen öznitelik alt kümeleri için, sınıflandırma algoritmaları çalıştırılarak başarı sonuçları karşılaştırılmıştır.

2.4. Sınıflandırma Algoritmaları

Bu çalışmada, sınıflandırma işlemlerinin gerçekleştirilmesi için açık kaynak kodlu olup, Python programlama dili ile çalışan scikit-learn makine öğrenmesi kütüphanesi kullanılmıştır. Scikit-learn makine öğrenmesi kütüphanesi son zamanlarda popülerliği artmakta ve literatürde kullanımı artmaktadır [16,17]. Tüm sınıflandırma işlemleri Python'un 3.7.3 sürümü ve scikit-learn kütüphanesinin 0.20.3 sürümü ile gerçekleştirilmiştir. Kullanılan sınıflandırma yöntemleri Tablo 2'de verilmiştir.

AdaBoostClassifier: Adaptive Boosting algoritması, her bir öznelik için zayıf bir sınıflandırıcı oluşturarak, bu zayıf sınıflandırıcıların birleşiminden ortaya güçlü bir sınıflandırıcının çıkarılması esasına dayanan bir algoritmadır. En iyi sınıflandırma işlemi yapan zayıf sınıflandırıcılar tespit edilir ve iyileştirme işlemleri yapılır. Son olarak en iyi sınıflandırma yapan zayıf sınıflandırıcılar birleştirilerek sınıflandırma işlemi gerçekleştirilir.

MLPClassifier: Çok Katmanlı Algılayıcı algoritması, beyindeki öğrenme mekanizmasından esinlenilerek geliştirmiş ve literatürde sıklıkla kullanılan bir sınıflandırma algoritmasıdır. Nöron yapısının simüle edildiği bu yöntemde, girdilere karşılık çıktığı en iyi şekilde sınıflandırabilecek bir ağ yapısı oluşturulmaktadır. Nöronlar arası geçişlerde kullanılan ağırlıkların optimum şekilde belirlenmesi ile sınıflandırma işlemleri gerçekleştirilebilmektedir.

Tablo 1. Sınıflandırma işlemi için çıkarılan öznelikler

Renk Uzayı	No	Öznelik	No	Öznelik	No	Öznelik
RGB Renk Uzayı R Değeri	1	MinR	2	MaxR	3	OrtR
RGB Renk Uzayı G Değeri	4	MinG	5	MaxG	6	OrtG
RGB Renk Uzayı B Değeri	7	MinB	8	MaxB	9	OrtB
HSV Renk Uzayı H Değeri	10	MinH	11	MaxH	12	OrtH
HSV Renk Uzayı S Değeri	13	MinS	14	MaxS	15	OrtS
HSV Renk Uzayı V Değeri	16	MinV	17	MaxV	18	OrtV
XYZ Renk Uzayı X Değeri	19	MinX	20	MaxX	21	OrtX
XYZ Renk Uzayı Y Değeri	22	MinY	23	MaxY	24	OrtY
XYZ Renk Uzayı Z Değeri	25	MinZ	26	MaxZ	27	OrtZ
YCbCr Renk Uzayı Y Değeri	28	Min_Y	29	Max_Y	30	Ort_Y
YCbCr Renk Uzayı Cb Değeri	31	MinCb	32	MaxCb	33	OrtCb
YCbCr Renk Uzayı Cr Değeri	34	MinCr	35	MaxCr	36	OrtCr
Gri Renk Uzayı	37	MinGri	38	MaxGri	39	OrtGri
Genişlik/Yükseklik Oranı	40	Oran				

SVC: Destek Vektör Makinesi algoritması, iki ya da daha fazla sınıfı birbirinden ayıran hiper-düzlemin belirlenmesine dayalı bir sınıflandırma algoritmasıdır. Sınıfları birbirinden ayırmak için sonsuz adet düzlem belirlenebilmektedir. Destek Vektör Makinesi algoritmasında amaç, sınıfları birbirinden ayıran ve sınıf örneklerinin hiper-düzleme olan mesafelerinin maksimum olduğu bir hiper-düzlemin belirlenmesidir.

DecisionTreeClassifier: Karar Ağaçları algoritması, sınıflandırma işlemi yaparken bir ağaç yapısı oluşturmaya dayanır. Oluşturulan ağaçların yaprakları sınıf etiketlerini temsil ederken, kökten yapraklara kadar giden yollar özneliklere göre dallanmaları temsil etmektedir.

KNeighborsClassifier: k-En Yakın Komşuluk algoritması, veri setindeki girdiler ve çıktıların dikkate alındığı bir danışmanlı öğrenme algoritmasıdır. Bu algoritmada, k adet en yakın komşuluğa bakılarak, gelen örneğin hangi sınıfa ait olduğunu tespit edilmektedir. Hızlı ve başarılı sonuçlar veren bir sınıflandırma yöntemidir.

GaussianNB: Olasılık tabanlı çalışan bir sınıflandırma yöntemidir. Bu yöntemde girdiler birbirlerinden bağımsız olarak düşünülerek, her bir girdinin çıktığı tahmin etme olasılığı hesaplanır. Eğitim verilerinden faydalanarak, yeni gelen girdilerin hangi sınıfa ait olduğunu hesaplamaya dayalı bir istatistiksel yöntemdir.

RandomForestClassifier: Rasgele Orman algoritması, birden fazla karar ağacının kullanıldığı ve mantık olarak Karar Ağaçları algoritması ile aynı şekilde çalışan bir algoritmadır. Birden fazla karar ağacı oluşturularak sınıflandırma başarısının artırılması amaçlanmaktadır.

Tablo 2. Sınıflandırma işlemi için kullanılan yöntemler

Scikit-learn Kütüphanesi Sınıflandırıcı	Açıklama
sklearn.neighbors.KNeighborsClassifier	<i>K</i> en yakın komşu sınıflandırma yöntemi
sklearn.svm.SVC	Destek vektör makinesi sınıflandırma yöntemi
sklearn.tree.DecisionTreeClassifier	Karar ağacı sınıflandırma yöntemi
sklearn.ensemble.RandomForestClassifier	Rasgele orman sınıflandırma yöntemi
sklearn.ensemble.AdaBoostClassifier	Adaptive Boosting sınıflandırma yöntemi
sklearn.naive_Bayes.GaussianNB	Gaussian Naive Bayes sınıflandırma yöntemi
sklearn.neural_network.MLPClassifier	Çok katmanlı algılayıcı sınıflandırma yöntemi

2.5. Karşılaştırma Ölçütleri

Bu çalışmada sınıflandırma başarısının ölçümünde hata matrisi kullanılmış ve toplamda beş farklı ölçüt kullanılmıştır. Hata matrisi sınıf etiketi kadar satır ve sütundan oluşan ve sınıflandırma başarısı hakkında bilgiler veren bir matristir. Tablo 3'te kullanılan hata matrisi verilmiştir.

Tablo 3. Kullanılan hata matrisi

		Tahmin Sınıfı	
		0	1
Gerçek Sınıf	0	Doğru Negatif (DN)	Yanlış Pozitif (YP)
	1	Yanlış Negatif (YN)	Doğru Pozitif (DP)

Tablo 3'te belirtilen parametreler, sınıflandırma hakkında bilgiler vermektedir. DN değeri, gerçekte acı sınıfa ait olan ve tahmin sonucu da acı olan örnek sayısını, YN değeri, gerçekte acı sınıfa ait olan ama tahmin sonucu tatlı olarak sınıflandırılan örnek sayısını, YP değeri, gerçekte tatlı sınıfta olan ve acı olarak tahmin edilen örnek sayısını, son olarak DP değeri ise, gerçekte tatlı sınıfta olan ve tahmin sonucunda da tatlı olarak sınıflandırılan örnek sayısını ifade etmektedir.

Doğruluk ("accuracy"): Bu ölçüt, doğru olarak sınıflandırılmış örnek sayısının toplam örnek sayısına olan oranını ifade etmektedir. Denklem 1'de doğruluk hesabı verilmiştir.

$$D = \frac{DN + DP}{DN + YN + YP + DP} \quad (1)$$

Kesinlik ("precision"): Kesinlik ölçütü, tahmin edilen pozitif değerler içinde doğru olarak sınıflandırılmış örnek sayısının, toplam tahmin edilen pozitif örnek sayısına olan oranını ifade etmektedir. Denklem 2'de kesinlik hesabı verilmiştir.

$$K = \frac{DP}{YP + DP} \quad (2)$$

Geriçağırım ("recall"): Geriçağırım ölçütü, gerçek pozitif değerler içinde doğru olarak sınıflandırılmış örnek sayısının, tüm gerçek pozitif örnek sayısına olan oranını ifade etmektedir. Denklem 3'te geriçağırım ölçüt hesabı verilmiştir.

$$G = \frac{DP}{YN + DP} \quad (3)$$

F₁ ölçütü ("F₁ score"): Kesinlik ve geriçağırım ölçütlerinin harmonik ortalamasını ifade eden ölçüttür. Denklem 4'te ifade edilmiştir.

$$2 * \frac{K * G}{K + G} \quad (4)$$

Ortalama Mutlak Hata-OMH (“*Mean Absolute Error*”): OMH ölçütü, gerçek değerler ile tahmin edilen değerler arasındaki farkların toplamının ortalaması olarak ifade edilmektedir. Denklem 5’te OMH hesabı verilmiştir.

$$OMH = \frac{1}{n} \sum_{t=1}^n |y_t - y'_t| \quad (5)$$

3. Bulgular ve Tartışma

Bu bölümde, sınıflandırma algoritmaları ile veri seti üzerinde çalışmalar yapılmıştır. Öznitelik seçimi aşamasında farklı sayılarda öznitelik alt kümeleri oluşturularak sonuçlara olan etkisi incelenmiştir. Ayrıca kullanılan tüm yöntemler için hiper-parametre ayarlaması yapılarak, her bir yöntemden en iyi sonucu elde edebilecek parametreler belirlenerek başarı karşılaştırılması yapılmıştır.

Eğitim ve test verilerinin belirlenmesi, tüm yöntemlerde aynı şekilde olmak üzere “k katlamalı çapraz doğrulama” tekniği kullanılarak gerçekleştirilmiştir. Çapraz doğrulama tekniğinde k sayısı 10 olarak belirlenmiştir. Sınıflandırma işlemi yapılırken üç farklı sayıda öznitelik alt kümesi oluşturulmuştur. Sırasıyla 4, 10 ve 15 adet öznitelikten oluşan üç farklı sınıflandırma işlemi gerçekleştirilmiştir. Bu öznitelik sayıları, Rasgele Orman algoritmasının diğer algoritmalarından daha iyi olduğunu gösterdiği için seçilmiştir. Toplamda 40 adet öznitelik bulunan veri setinde düşük sayıda öznitelik ile başlayarak sınıflandırma işlemleri gerçekleştirilmiştir. Öznitelik seçim işlemi, sınıflandırma işlemine etkisi bakımından özniteliklerin güçlüden zayıfa doğru ilk 10 özniteliğin sıralaması Tablo 4’te verilmiştir. İlk sırada ise YCbCr Renk Uzayı S Değerinin MinCb özniteliği yer almaktadır. Renk bakımından sınıflandırma işlemi YCbCr Renk Uzayının önemi sonuçlardan anlaşılmaktadır. 40 nolu öznitelik yani kayısı iç çekirdeğinin genişlik/yükseklik oranı ikinci sırada yer almaktadır. Elde edilen sonuçlar da boyutun önemli bir faktör olduğunu desteklemektedir.

Tablo 4. En güçlü ilk 10 özniteliğin sıralaması

Sıra No	Öznitelik No	Öznitelik Adı
1	31	MinCb
2	40	Oran
3	6	OrtG
4	30	Ort_Y
5	39	OrtGri
6	24	OrtY
7	21	OrtX
8	18	OrtV
9	3	OrtR
10	36	OrtCr

Tablo 5’te dört adet özniteliğin seçildiği sınıflandırma işlemi sonuçları verilmiştir. Seçilen öznitelik numaralarının hangi özniteliklere karşılık geldiği Tablo 1’den bakılabilir. Tablo 6 ve Tablo 7’de sırasıyla 10 ve 15 adet öznitelik kullanılarak elde edilen sonuçlar gösterilmiştir.

Tablo 5. 6, 30, 31 ve 40 nolu öznitelikler kullanılarak elde edilen sonuçlar

Sınıflandırıcı	Doğruluk	Kesinlik	Geriçağırım	F ₁ ölçütü	OMH
KNeighborsClassifier	0,89	0,95	0,78	0,85	0,11
SVC	0,92	0,95	0,86	0,90	0,08
DecisionTreeClassifier	0,94	0,96	0,90	0,93	0,06
RandomForestClassifier	0,96	0,97	0,93	0,95	0,04
AdaBoostClassifier	0,90	0,89	0,87	0,88	0,10
GaussianNB	0,90	0,90	0,85	0,87	0,10
MLPClassifier	0,91	0,93	0,85	0,89	0,09

Tablo 6. 3, 6, 18, 21, 24, 30, 31, 36, 39 ve 40 nolu öznelikler kullanılarak elde edilen sonuçlar

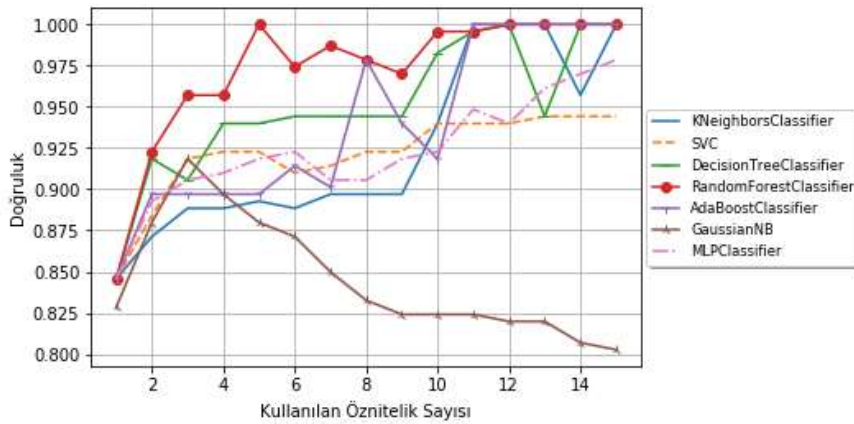
Sınıflandırıcı	Doğruluk	Kesinlik	Geriçağırım	F ₁ ölçütü	OMH
KNeighborsClassifier	0,94	0,96	0,90	0,93	0,06
SVC	0,94	0,96	0,90	0,93	0,06
DecisionTreeClassifier	0,98	1,00	0,96	0,98	0,02
RandomForestClassifier	1,00	1,00	0,99	0,99	0,00
AdaBoostClassifier	0,92	0,90	0,91	0,90	0,08
GaussianNB	0,82	0,84	0,72	0,78	0,18
MLPClassifier	0,92	0,92	0,90	0,91	0,08

Tablo 5, 6 ve 7 incelendiğinde en başarılı sınıflandırıcının Rasgele Orman algoritmasının olduğu görülmektedir. Dört öznelik kullanılarak yapılan sınıflandırma sonucunda Rasgele Orman algoritması 0.96'lık bir başarı oranı ile ilk sırada yer almaktadır. Öznelik sayısı artırıldıkça diğer sınıflandırıcılarında başarı oranının arttığı görülmektedir. Ancak daha az sayıda öznelik kullanarak daha iyi sonuç verebildiği için Rasgele Orman algoritması, bu çalışmada ele alınan problem için en başarılı yöntemdir.

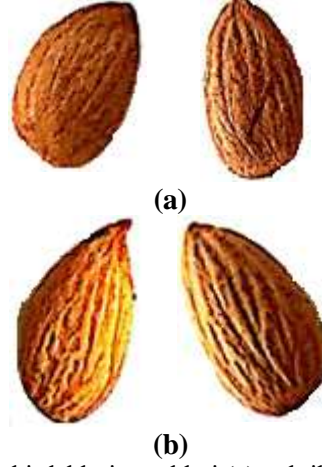
Şekil 2'de kullanılan sınıflandırma yöntemlerinin örnek bir çalıştırılma sonucunda, seçilen öznelik sayısına göre elde ettikleri başarılar verilmiştir. Yöntemlerin başarı karşılaştırmaları Şekil 2'den daha detaylı olarak görülebilmektedir. %100'lük bir başarı oranına ilk ulaşan yöntemin Rasgele Orman algoritmasının olduğu ve bu başarı için beş adet öznelik kullanmasının yeterli olduğu görülmektedir. Rasgele Orman algoritmasının performansının her durumda diğer algoritmalarından daha iyi olduğu anlaşılmaktadır. Adaptive Boosting, Karar Ağacı, k En Yakın Komşu algoritmalarının, Rasgele Orman algoritmasına ancak 10'dan fazla özneliği kullandığı durumda yetişebildiği de yine Şekil 2'den anlaşılmaktadır. Ayrıca kullanılan sınıflandırma yöntemleri içinde sadece Gaussian Naive Bayes sınıflandırıcının performansı öznelik sayısı arttıkça düşmekte iken diğer tüm yöntemlerin performansı artarak devam etmektedir.

Tablo 7. 3, 6, 9, 15, 18, 21, 22, 24, 27, 30, 31, 36, 37, 39 ve 40 nolu öznelikler kullanılarak elde edilen sonuçlar

Sınıflandırıcı	Doğruluk	Kesinlik	Geriçağırım	F ₁ ölçütü	OMH
KNeighborsClassifier	1,00	1,00	1,00	1,00	0,00
SVC	0,94	0,97	0,90	0,93	0,06
DecisionTreeClassifier	1,00	1,00	1,00	1,00	0,00
RandomForestClassifier	1,00	1,00	1,00	1,00	0,00
AdaBoostClassifier	1,00	1,00	1,00	1,00	0,00
GaussianNB	0,80	0,84	0,65	0,74	0,20
MLPClassifier	0,98	0,97	0,98	0,97	0,02

**Şekil 2.** Kullanılan öznelik sayısına göre algoritmaların karşılaştırılması

Şekil 3'te ise sınıflandırıcılar tarafından yanlış etiketlenen örnek kayısı iç çekirdekleri verilmiştir. Yanlış sınıflandırılan örnekler incelendiğinde boyut ve doku olarak sınıflar arası benzer olan örneklerin karıştırıldığı anlaşılmaktadır. Bu karışıklıkların engellenebilmesi için öznitelik çeşitliliğinin daha fazla artırılması gerektiği anlaşılmaktadır.



Şekil 3. Yanlış sınıflandırılan kayısı iç çekirdekleri örnekleri (a) tatlı iken acı olarak sınıflandırılan örnekler (b) acı iken tatlı olarak sınıflandırılan örnekler

4. Sonuç ve Öneriler

Bu çalışmada kayısı iç çekirdeklerinin acı ve tatlı olarak görüntü işleme ve makine öğrenmesi algoritmaları kullanarak sınıflandırılması işlemi yapılmıştır. Yedi farklı makine öğrenmesi algoritması kullanarak elde edilen sonuçlar karşılaştırılmış ve Rasgele Orman algoritmasının en başarılı sınıflandırıcı olduğu sonucuna varılmıştır. Az sayıda öznitelik ile en iyi sonucu veren Rasgele Orman algoritması yeterli sayıda öznitelik sayısına ulaşıncaya kadar yüzde yüz bir başarı ile sınıflandırma işlemi yapabilmektedir. Rasgele Orman algoritmasına alternatif olarak k en yakın komşu algoritması, karar ağacı algoritması ve Adaptive Boosting algoritması düşünülebilmektedir. Fakat en az sayıda öznitelik kullanılarak, en yüksek başarıyı Rasgele Orman algoritması elde edebilmektedir.

Gelecekteki çalışmalarda, kayısı iç çekirdeğinin derin öğrenme algoritmaları ile sınıflandırılması amaçlanmaktadır. Böylece öznitelik oluşturmakla kullanıcılar zaman kaybetmeden doğrudan resimleri kullanarak öznitelikler otomatik olarak derin öğrenme ağ yapısı ile oluşturulacak ve sınıflandırma işlemi gerçekleştirilecektir.

Yazarların Katkısı

Makalede tüm katkı şahsıma aittir.

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.

Kaynaklar

- [1] Öztürk N., Ulusoy M.R., Erkiş L., Bayhan S.Ö. 2004. Malatya ili kayısı bahçelerinde saptanan zararlılar ile avcı türler. Bitki Koruma Bülteni, 44 (1-4): 1-13.

- [2] Wen X., Jin F., Regenstein J.M., Wang F. 2018. Transglutaminase induced gels using bitter apricot kernel protein: Chemical, textural and release properties. *Food Bioscience*, 26: 15-22.
- [3] Kaya A., Okur M., Üstyoğ L., Temel H., Çaksen H. 2012. Kayısı çekirdeği yeme sonrası akut siyanür zehirlenme olgusu. *Türk Pediatri Arşivi*, 47 (2): 141-142.
- [4] Özcan M., Özcan F.B.T., Yaşartekin Y., Yavuz H., Sarıcı S.Ü. 2017. Kayısı çekirdeğine bağlı akut siyanür zehirlenmesi, *Cukurova Medical Journal*, 42(3): 600–601.
- [5] Karhan M., Oktay M.O., Karhan Z., Demir H. 2011. Morfolojik Görüntü İşleme Yöntemleri ile Kayıslarda Yaprak Delen (Çil) Hastalığı Sonucu Oluşan Lekelerin Tespiti. 6th International Advanced Technologies Symposium (IATS'11), 16–18 May, Elazığ, pp. 172-176.
- [6] Khojastehnazhand M., Mohammadi V., Minaei S. 2019. Maturity detection and volume estimation of apricot using image processing technique. *Scientia Horticulturae*, 251: 247-251.
- [7] Hussain Hassan N.M., Nashat A.A. 2019. New effective techniques for automatic detection and classification of external olive fruits defects based on image processing techniques. *Multidimensional Systems and Signal Processing*, 30 (2): 571-589.
- [8] Alam M.N., Pineda I., Lim J.G., Gwon O. 2018. Apple Defects Detection Using Principal Component Features of Multispectral Reflectance Imaging. *Science of Advanced Materials*, 10 (7): 1051-1062.
- [9] Ye D., Sun L., Tan W., Che W., Yang M. 2018. Detecting and classifying minor bruised potato based on hyperspectral imaging. *Chemometrics and Intelligent Laboratory Systems*, 177: 129-139.
- [10] Wan P., Toudeshki A., Tan H., Ehsani R. 2018. A methodology for fresh tomato maturity detection using computer vision. *Computers and electronics in agriculture*. 146: 43-50.
- [11] Nasirahmadi A., Ashtiani S-H.M. 2017. Bag-of-Feature model for sweet and bitter almond classification. *Biosystems Engineering*. 156. 51-60.
- [12] Thong N.D., Thinh N.T., Cong H.T. 2019. Mango Sorting Mechanical System Uses Machine Vision and Artificial Intelligence. *International Journal of Engineering and Technology*, 11 (5).
- [13] Luan Z., Li C., Ding S., Wei M., Yang Y. 2020. Sunflower seed sorting based on Convolutional Neural Network. In Eleventh International Conference on Graphics and Image Processing (ICGIP 2019), 11373: 113731K.
- [14] Zeng X., Miao Y., Ubaid S., Gao X., Zhuang S. 2020. Detection and classification of bruises of pears based on thermal images. *Postharvest Biology and Technology*, 161: 111090.
- [15] Zhao B., Wang Y., Fu J., Zhao R., Li Y., Dong X., Lv C., Jiang H. 2020. Online Measuring and Size Sorting for Perillae Based on Machine Vision. *Journal of Sensors*, 2020 (Özel Sayı): 1-8.
- [16] Kalaycı T.E. 2018. Comparison of Machine Learning Techniques for Classification of Phishing Web Sites. *Pamukkale University Journal of Engineering Sciences*, 24 (5): 870-878.
- [17] Aydın C. 2018. Makine Öğrenmesi Algoritmaları Kullanılarak İtfaiye İstasyonu İhtiyacının Sınıflandırılması. *European Journal of Science and Technology*, 14: 169-175.