



## K-Star, Rastgele Orman ve Karar Ağacı (C4.5) Sınıflandırma Algoritmaları ile Domatesin Renk Olgunluğu Üzerinde Bazı Mekanik Özelliklerin Etkisinin Belirlenmesi

Hande Küçükönder<sup>1\*</sup>, Kubilay Kazım Vursavuş<sup>2</sup>, Fatih Üçkardes<sup>3</sup>

<sup>1</sup>Bartın Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, 74100 Bartın, Türkiye

<sup>2</sup>Çukurova Üniversitesi, Ziraat Fakültesi, Tarım Makinaları Bölümü, 01330 Balcalı-Sarıçam/Adana, Türkiye

<sup>3</sup>Adıyaman Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi, 02040 Adıyaman, Türkiye

### MAKALE BİLGİSİ

Geliş 28 Ekim 2014  
Kabul 12 Ocak 2015  
Çevrimiçi baskı, ISSN: 2148-127X

#### Anahtar Kelimeler:

Domates  
Mekanik özellik  
Renk ölçüm  
K-Star  
Random Forest  
Karar Ağacı (C4.5)

#### \* Sorumlu Yazar:

E-mail: hkucukonder@bartin.edu.tr

### Ö Z E T

Bu çalışma, domatesin renk olgunluğu üzerine kabuk yırtılma noktasındaki kuvvet, kabuk yırtılma noktasındaki enerji ve kabuk yırtılma sertliği gibi mekanik özelliklerin etkisinin veri madenciliği yönteminin danışmanlı öğrenme algoritmaları ile belirlenmesi amacıyla yürütülmüştür. Araştırmada toplam 88 adet domates örneği kullanılmış ve her domateste 4 farklı ekvatorial bölgede renk ölçümü gerçekleştirilerek toplamda 352 adet renk ölçümü kullanılmıştır. Mekanik özelliklere göre yapılan sınıflandırma işlemlerinde veri madenciliğinin K-Star, Random Forest ve Karar Ağacı (C4.5) algoritmalarından yararlanılmış ve oluşan sınıflandırma modellerinin karşılaştırmasında hata varyans kriterlerinden, Hata kareler ortalamasının karekökü (Root Mean Square Error-RMSE), Ortalama mutlak hata (Mean absolute error-MAE), Bağlı karesel hatanın karekökü (Root relative squared error-RRSE) ile Göreli mutlak hata (Relative absolute error – RAE) değerlerinin düşük, sınıflandırma doğruluk oranının ise yüksek olması dikkate alınmıştır. Yapılan karşılaştırma sonucunda ise örnek tabanlı K-Star algoritmasına göre oluşan [MAE: 0,004, RMSE: 0,006, %RAE: 1,73 ve %RRSE: 1,70] sınıflandırma modelinin diğerlerine nazaran daha iyi sınıflandırıcı olduğu tespit edilmiştir. K-Star algoritmasına göre yapılan sınıflandırma ile domateste olgunluk derecesi üzerinde kabuk yırtılma noktasındaki kuvvet ile kabuk yırtılma sertliği özelliklerinin yeşil, açık kırmızı ve renk dönüşüm dönemlerinde etkisinin önemsiz diğer dönemlerde önemli kabuk yırtılma noktasındaki enerjinin ise yalnızca pembe ve renk dönüşüm safhalarında etkili diğer renk oluşum safhalarında ise önemsiz olduğu bulunmuştur.

Turkish Journal Of Agriculture - Food Science And Technology, 3(5): 300-306, 2015

## Determining The Effect of Some Mechanical Properties on Color Maturity of Tomato With K-Star, Random Forest and Decision Tree (C4.5) Classification Algorithms

### ARTICLE INFO

#### Article history:

Received 28 October 2014  
Accepted 12 January 2015  
Available online, ISSN: 2148-127X

#### Keywords:

Tomato  
Mechanical properties  
Color measurement  
K-Star  
Random Forest  
Decision Tree (C4.5)

#### \* Corresponding Author:

E-mail: hkucukonder@bartin.edu.tr

### ABSTRACT

This study was conducted in order to determine the effect of the mechanical properties such as maximum force at the skin rupture point, energy at the skin rupture point and the skin firmness on color maturity of tomato by supervised learning algorithms of data mining. In the present study, a total of 88 tomato samples were used, and color measurements for each tomato in 4 different equatorial regions were performed and a total of 352 color measurement units were used. In the classification processes performed according to these mechanical properties, K-Star, Random Forest and Decision Tree (C4.5) algorithms of data mining were utilized, and in the comparison of comprising classification models, Root Mean Square Error (RMSE), Mean absolute error (MAE), Root relative squared error (RRSE) and Relative absolute error (RAE) values, which are some of the criteria of error variance, were considered to be low, while the classification accuracy rate was considered to be high. As a result of the comparison made, the classification model formed according to K-Star instance-based algorithm [MAE: 0.004, RMSE: 0.006, %RAE: 1.73, %RRSE: 1.70] has been found to be a better classifier compared to the others. With the classification made according to K-Star algorithm, the maximum force at the skin rupture point on the degree of maturity of tomato and the skin firmness were found to be green, light red, and their effects are non-significant during the color conversion periods, and found significant during other periods while the energy at the skin rupture point is only pink and has been to be significant during the color conversion stages and non-significant during other stages.

## Giriş

Dünyada ve ülkemizde tarımsal üretimin yeterli ve kaliteli bir şekilde yapılması özellikle de nüfus ihtiyaçlarının karşılanması açısından büyük önem taşımaktadır. Tarımsal ürünlerin belirli bir standarda uyma zorunluluğunun olması, bir takım fiziksel özelliklerin yanı sıra zedelenmemiş ve sağlam olması da hem ekonomik anlamda hem de pazar payı açısından etkili olan faktörler arasındadır (Yurtlu ve Erdoğan, 2005). Özellikle de bazı statik ve dinamik dış kuvvetlerin etkisiyle meydana gelebilecek mekanik zedelenmeler (Vursavuş ve Özgüven, 2007) genel olarak en fazla hasat esnasında ve sonrasında çarpmayla oluşan kuvvetlerin ve aşırı deformasyonun etkisiyle ortaya çıkmaktadır (Eraltan, 2005).

Tarımsal ürünlerden yumuşak yapılı olanlar (domates, seftali ve kivi gibi) mekanik zedelenmelere karşı daha duyarlı olduklarından dolayı bu gibi ürünlerin pazara iletilmesine kadar geçen sürede uygulanan işlemler sırasında oluşan bir takım zedelenmeler hem pazar değerini düşürmekte hem de ürünlerin bir takım bozulmalara maruz kalmalarına neden olmaktadır (Kara ve Turgut, 1988). Bu durum aynı zamanda ürünlerin mekanik özelliklerinin bilinmesinin ve bu özelliklere göre yapılacak olan ürün sınıflandırmasının da önemini vurgulamakta olup meydana gelebilecek olası zararın azaltılmasında etkili olduğunu da işaret etmektedir.

Günümüzde gelişen teknolojiyle birlikte, araştırmacıların tarımsal ürünlerin kalite ve özelliklerine göre sınıflandırılmasında bazı matematiksel tahmin modellerini kullanması (lineer regresyon, doğrusal ayırım) ve buna yönelik olarak hazırlanan programlar ile zaman ve iş gücü bakımından büyük kayıpların oluşması engellenmektedir. Bu amaca yönelik olarak makine öğrenmesi gerçekleştiren algoritmalar ile özellikle de hızlı karar vermeyi gerektiren durumlar için son yıllarda yaygın kullanılan tekniklerden biri de Veri Madenciliği'dir. Veri madenciliği, endüstri, tıp, mühendislik ve pazarlama stratejisi gibi bir çok alanda tahmin ve sınıflandırma amacıyla yaygın bir şekilde kullanılmakta olup (Camdeviren ve ark., 2007; Lee ve ark., 2006) Tarımsal alanda kullanımı ile ilgili sınırlı sayıda çalışma bulunmaktadır. Sencan (2004), tarımsal alanlarda yetiştirilen bazı ürünler için, Jay ve ark. (2009), karmaşık ve homojen yeşil alan topluluklarını belirlemede, Er ve ark. (2013) elma çeşitlerinin rengine göre sınıflandırılmasında söz konusu yöntemin farklı algoritmalarından yararlanmışlar ve yüksek doğruluk oranlarına ulaştıklarını bildirmişlerdir.

Bu çalışmanın amacı, domates bitkisinin renk olgunluğu üzerine kabuk yırtılma noktasındaki kuvvet ( $F_{max}$ ), kabuk yırtılma noktasındaki enerji [ $0,5(F_{max}D_{max})$ ]:Nmm) ve kabuk yırtılma sertliği ( $F_{max}/D_{max}$ :N/mm) mekanik özelliklerin etkisinin veri madenciliği sınıflandırıcıları [K-Star, Rastgele Orman, Karar Ağacı (C4.5)] ile belirlenmesidir.

## Materyal ve Yöntem

### Deneyel tasarım ve renk ölçümleri

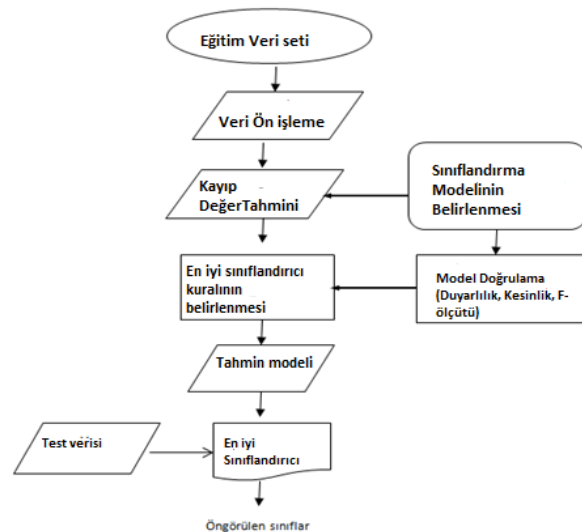
Denemeler de 2014 yılı Şubat ayında elle hasat edilmiş olan Bandita F1 salkım sırik domates çeşidi kullanılmıştır. Toplam 88 adet domates örneği araştırma

süresince test edilmiş ve her domates örneğinde 4 farklı ekvatorial bölgede renk ölçümü gerçekleştirilmiştir. Böylece toplamda 352 adet renk ölçümü araştırma süresince kullanılmıştır. Domates rengi Minolta CR-400 model renk ölçüm cihazı kullanılarak  $L^*$ ,  $a^*$ ,  $b^*$  üç nokta ölçüm yöntemiyle belirlenmiştir. Denemelerde kullanılan domates örneklerine ait fiziksel özellikleri Çizelge 1'de verilmiştir.

Araştırmada domateslerin kırmızılık değerleri ve olgunluk sınıflandırmaları ise Batu (2004) adlı araştırmacı tarafından önerilen  $a^*/b^*$  renk uzayı değer aralıkları kullanılarak saptanmıştır (Çizelge 2). Böylece 6 farklı olgunluk düzeyi elde edilmiştir. Minolta renk ölçüm cihazında  $a^*$  değeri domateslerin kırmızılık derecesini,  $b^*$  değeri de yeşillik derecesini belirtmektedir. Denemelerde kullanılan domates örneklerine ait renk değerleri Çizelge 3'de verilmiştir.

Domateslerin sertlik ölçümlerinde Lloyd Üniversal biyolojik materyal test cihazı kullanılmıştır. Bu ölçümlerde 4 mm çaplı düz batıcı uç kullanılmış ve 10 mm  $\text{min}^{-1}$  lik yükleme hızında çalışılmıştır (ASAE Standards, 2009). Domates örneklerine ait kuvvet deformasyon grafikleri oluşturulurken batıcı uca domates örneği içerisinde 10 mm yol aldırılmıştır. Kabuk yırtılma noktasındaki kuvvet ( $F_{max}$ ; RF) çizelgelerde belirtildiği gibi RF ile deformasyon ( $D_{max}$ ) değerleri kullanılarak kabuk yırtılma sertliği ise F ile çizelgede belirtildiği F şeklinde  $F:(F_{max}/D_{max}:\text{Nmm}^{-1})$  kabuk yırtılma enerjisi RE şeklinde çizelgedeki gibi RE simgesi şeklinde  $RE:[0,5(F_{max}D_{max}):\text{N mm}]$  simgelenendirilebilir. Mekanik özelliklerinin hesaplanmalarda, kopma (kırılma) absorbe enerjisi genelde kuvvet deformasyon eğrisi altındaki alan olarak verilir. Formül olarak tam olarak karşılımsa da  $AE=(F.D)/2$  şeklinde verilir (Braga et al., 1999). Böylece kabuk yırtılma kuvveti, kabuk sertliği ve kabuk yırtılma enerjisi parametreleri istatistiksel hesaplamalarda kullanılmıştır.

Domatesin renk olgunluğu üzerine mekanik özelliklerin etkisini belirlemek amacıyla kullanılan algoritmaların veri setine uygulanması için tasarlanan sistem Şekil 1'deki gibi dizayn edilmiştir.



Şekil 1 Domateste renk olgunluğunu belirlemek için önerilen sistemin çerçevesi

*İstatistiksel analiz*

*K-Star Algoritması:* Veri madenciliğinde kullanılan öğrenme algoritmalarından biri olan bu algoritmada amaç, test veri setinde yer alan özneliği bilinmeyen bir örneğin, veri tabanı içerisinde daha önceden sınıflandırılmış fakat ortaya çıkmamış gizli kalmış olan eğitim veri setindeki örnekler ile karşılaştırılması esasına göre sınıflandırma yapmaktır (Aha ve ark.1991). Çalışmada kullanılan K-star sınıflandırıcısı Çölkesen (2010)'nin bildirdiğine göre aşağıda yer alan denklemler aracılığı ile hesaplanmaktadır.

$$P^*\left(\frac{b}{a}\right) = \sum_{t \in P_{jt(a)=b}} P(t) \quad (1)$$

K-Star fonksiyonu ise,

$$K^*\left(\frac{b}{a}\right) = -\log_2 P^*\left(\frac{b}{a}\right) \quad (2)$$

şeklinde hesaplanmaktadır. Bu eşitlikte yer alan  $P^*$ , Olasılık fonksiyonu;  $t$ , dönüşümleri göstermektedir. Rastgele bir  $P^*$  fonksiyonu içinde etkili olan örnek sayısı

$$n_0 \leq \frac{\sum_b P^*(b/a)^2}{\sum_b P^*(b/a)} \leq N \quad (3)$$

eşitliği ile bulunmaktadır. Buradaki,  $N$ : eğitim örneklerinin tamamının sayısını,  $n_0$ : eğitim örnekleri içerisinde  $a$  örneğine en kısa mesafedeki örneklerin sayısını göstermektedir (Çölkesen, 2010).

*Rastgele Orman (Random Forest):* Rastgele orman (RO) yöntemi, karar ağacı sınıflandırıcılarının bileşimini içeren bir yöntemdir. Karar ağacında olduğu gibi bu yöntemde de dallanma kriterlerinin belirlenmesinde en yaygın kullanılan teknik Gini indeksidir. Bu algoritmanın işleyişinde 2 farklı parametre esas alınmakta olup, bunlar sırasıyla her bir düğümde kullanılan değişken sayısı ve geliştirilecek olan ağaç sayısı'dır. RO ile ağaç üretmek için CART (sınıflandırma ve Regresyon ağacı) algoritması kriterlerinden Gini indeksi, Akar ve Güngör (2010)'ün bildirdiğine göre, aşağıda belirtilen denklem aracılığıyla hesaplanmaktadır.

$$\sum \sum j \neq i (f(C_i, T)/T) / (f(C_i, T)/T) \quad (4)$$

Bu eşitlikte,  $T$ , eğitim veri setini,  $C_i$ , rastgele seçilen bir domates renk ölçümünün ait olduğu sınıfı ve  $f(C_i, T)/T$ , seçilen örneğin  $C_i$  sınıfına ait olma olasılığını göstermektedir (Akar ve Güngör, 2010).

*Karar Ağacı (C4.5):* Karar ağaçları, ağaç yapısında anlaşılması ve yorumlanması kolay, verilerin hızlı bir şekilde işlenmesinin mümkün olduğu veri madenciliği sınıflandırıcılarındandır. Bu yöntemin önemli avantajları arasında özellikle sınıf yapıları daha önceden belirlenen kategorik veri yapılarında sınıfı bilinmeyen verilerin ait olduğu sınıfın tahmin edilmesi, ağaç şeklinde bilgiyi görsel olarak sunması ve değişkenler arasındaki ilişkinin kolaylıkla modellenmesi yer almaktadır (Biggs ve ark., 1991). Karar ağaçları, kök, düğüm ve dal yapısında yukarıdan aşağıya doğru bir ağaç görünümü olup bu

yapıda, kök ve her bir düğüm(node) bir soruyla, düğümlerden ayrılan dallar ise ilgili sorunun olası cevaplarını bildirecek şekilde bir dallanma sergilemektedir (Loh ve Shih, 1997). Her dal düğümü aynı zamanda bir tahmin sınıfı olarak da görülmektedir. Karar ağaçlarında, bilgi kazancı metodu (ID3, C5) bilgi oranı ve gini indeksi tekniklerinden yararlanılmakta olup bunlar içerisinde en yaygın olarak kullanılanı ise bilgi kazancı metodu olarak bilinmektedir. Kavzoğlu ve Çölkesen (2010)'nin bildirdiğine göre, bir veri setinde sınıflar  $C_1, C_2, \dots, C_n$  ve  $T$ 'de sınıf değerlerini göstermek üzere, bir sınıfa ait olasılık,  $P_i = (C_i/T)$  şeklinde olup sınıflara ait entropi değeri,

$$\text{Entropi}(T) = -\sum_{i=1}^n p_i \log_2(p_i) \quad (5)$$

şeklinde hesaplanmaktadır.  $T$  sınıf değerleri olmak üzere veri setinde belirlenen  $B$  öz nitelik değerine göre sınıf değerleri  $T_1, T_2, \dots, T_n$  olmak üzere alt kümeler bölünerek, elde edilecek olan kazanç ise

$$\text{Kazanç}(B, T) = \text{Entropi}(T) - \sum_{i=1}^n \frac{|T_i|}{|T|} \text{Entropi}(T_i) \quad (6)$$

eşitliği ile hesaplanmaktadır.  $B$  öz niteliği değerinin belirlenmesi içinde bölünme bilgisi ve kazanç oranı ise aşağıdaki eşitlikler ile belirlenmektedir (Kavzoğlu ve Çölkesen, 2010).

$$\text{Bölünme bilgisi}(B) = -\sum_{i=1}^k \frac{|T_i|}{|T|} \log_2\left(\frac{|T_i|}{|T|}\right) \quad (7)$$

$$\text{Kazanç oranı} = \frac{\text{Kazanç}(B, T)}{\text{Bölünme bilgisi}(B)} \quad (8)$$

*Doğruluk ölçütleri:* Model başarısı değerlendirilirken kullanılan ölçütler arasında hata oranı, kesinlik, duyarlılık ve F-ölçütü yer almaktadır. Söz konusu bu ölçütlerin hesaplanmasında Çizelge 4'de yer alan düzensizlik matrisinden yararlanılmaktadır (Çoşkun ve Baykal, 2011). Çizelgede yer alan satırlar test kümesindeki örneklere ait gerçek değerleri, sütunlar ise modelin tahminlemesini ifade etmekte olup Çoşkun ve Baykal (2011)'in bildirdiğine göre hesaplamalarda kullanılan denklem eşitlikleri ve açıklamaları aşağıda verilmiştir.

a: DP (Doğru Pozitif)

b: HN (Hatalı Negatif)

c: HP (Hatalı Pozitif)

d: DN (Doğru Negatif)

Doğruluk =  $TP + TN / TP + TN + FP + FN$

Hata oranı =  $TP + FN / TP + TN + FP + FN$

Kesinlik =  $TP / TP + FN$

Duyarlılık =  $TP / TP + FN$

F Ölçütü =  $2 \times \text{Duyarlılık} \times \text{Kesinlik} / (\text{Duyarlılık} + \text{Kesinlik})$

Verilerin analizinde WEKA programından (Weka, 2013) yararlanılmış, algoritmalara göre oluşan modeller, hata varyans kriterlerinden hata kareler ortalamasının karakökü (Root Mean Square Error-RMSE), ortalama mutlak hata (Mean absolute error -MAE), bağıl karesel hatanın karakökü (Root relative squared error- RRSE), görel mutlak hata (Relative absolute error-RAE) kriterlerine göre karşılaştırılmıştır. Söz konusu eşitlikler ile eşitliklerde yer alan terimler aşağıda verilmiştir.

$$K = \frac{P_G - P_B}{1 - P_B} \quad (9)$$

$$P_G = \frac{\sum_i G_{ii}}{n} \quad (10)$$

$$P_B = \frac{\sum_i R_i C_i}{n^2} \quad (11)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |t_i - o_i| \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |t_i - o_i|^2} \quad (13)$$

$$RAE = \frac{\sum_{i=1}^n |t_i - o_i|}{\sum_{i=1}^n |t_i - \bar{t}|} \quad (14)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^n |t_i - o_i|^2}{\sum_{i=1}^n |t_i - \bar{t}|^2}} \quad (15)$$

$P_G$  ve  $P_B$  sırasıyla gözlenen ve beklenen olasılıkları,  $G_{ii}$ :i-inci satır ve i-inci sütundaki gözlenen frekansı,  $R_i$ :i-inci satırdaki toplam frekansı,  $C_i$ :i-inci sütundaki toplam frekansı ve n:toplam gözlem sayısını (Keskin,2004),  $\bar{t}$  hedef değerlerin ortalamasını,  $t_i$ :i-inci sınıfın tahmin değeri ve  $o_i$ : i-inci sınıfın gerçek değerini göstermektedir (Weka, 2013). Kappa istatistiğinin uyum derecesi ise; [ $<0$ ] kötü, [ $0,01-0,20$ ] zayıf, [ $0,21-0,40$ ] fena değil [ $0,41-0,60$ ] orta, [ $0,61-0,80$ ] iyi [ $0,81-0,92$ ] çok iyi, [ $0,93-1$ ] mükemmel şeklinde sınıflandırılmıştır (Boyacıoğlu ve Güneri, 2006).

Çizelge 1 Bandita F1 domates çeşidine ait bazı fiziksel özellikler

Özellikler	Ortalama Değerler ( $\pm$ )
Ekvatorial çap (mm)	60,49 $\pm$ 4,01
Kalınlık (mm)	59,72 $\pm$ 3,75
Yükseklik (mm)	48,37 $\pm$ 4,16
Kütle (g)	103,89 $\pm$ 17,02
Hacim (cm <sup>3</sup> )	92,03 $\pm$ 15,93
Küresellik (%)	88,78 $\pm$ 2,03
Yüzey Alanı (cm <sup>2</sup> )	90,66 $\pm$ 10,09

Çizelge 2 Renk ölçümleri

Renk safhası	Sınıfı	a*/b*
1	Yeşil olum dönemi	-0,59<a*/b* $\leq$ -0,47
2	Renk kırılma dönemi	-0,47<a*/b* $\leq$ -0,27
3	Renk dönüşüm dönemi	-0,27<a*/b* $\leq$ -0,08
4	Pembe olum	0,08<a*/b* $\leq$ -0,60
5	Açık kırmızı olum	0,60<a*/b* $\leq$ -0,95
6	Kırmızı olum	a*/b*>0,95

Çizelge 3 Denemelerde Kullanılan Domates Örneklerine Ait Ortalama Renk Değerleri

Özellikler	Ölçüm Değerleri
L*	42,44 $\pm$ 2,53
a*	8,17 $\pm$ 11,74
b*	18,96 $\pm$ 2,64
a*/b*	0,36 $\pm$ 0,58

Çizelge 4 Model başarı ölçütlerinin belirlendiği düzensizlik matrisi\*

Gerçek sınıf	Öngörülen sınıf	
	Sınıf 1	Sınıf 0
Sınıf1	a	b
Sınıf 0	c	d

\*(Coşkun ve Baykal , 2011)

## Bulgular

Sınıflandırmada, bağımsız değişken olarak kullanılan mekanik özellikler ile bağımlı değişken renk oluşum safhalarına ait tanımlayıcı istatistiklerinin değerleri Çizelge 5’de verilmiştir. Araştırmada, domates veri setine K-Star, C4.5 ve Rastgele Orman uygulanmış ve modellere ait hata performans kriterlerinin değerleri Çizelge 6 ’deki gibi bulunmuştur.

Sınıflandırmalara göre oluşturulan modellerin hata performans kriterleri, MAE: 0,004-0,13, RMSE: 0,006-0,26, %RAE: 1,73-50,81 ve %RRSE: 1,70-70,31 değer aralıklarında (Çizelge 6) ve model başarı ölçütlerinden Duyarlılık: 0,707-1; Kesinlik: 0,726-1; F ölçütü: 0,703-1 arasında değiştiği belirlenmiştir (Çizelge 7). Algoritmalar bu kriterler bakımından karşılaştırıldığında ise en küçük hata değerine sahip olan K-Star algoritmasının diğerlerine göre daha başarılı sınıflandırma yaptığı belirlenmiştir (Çizelge 6-7). K star algoritmasına göre toplam 352 adet domates renk ölçüm değerleri ait olduğu renk sınıflarına minimum hata ile %100 doğruluk oranıyla atanmıştır. Sınıflandırma sonucunda, 57 tane domatesin renk ölçüm değerlerine göre yeşil, 24 tanesinin kırılma, 93 tanesinin pembe, 67 tanesinin açık kırmızı, 35 adeti renk dönüşüm ve 76 tanesinin de kırmızı renkli olduğu tespit edilmiştir (Şekil 2a).

Domateslerin renk olgunluklarına göre sınıflara ayrıştırılmasında, kabuk yırtılma noktasındaki kuvvet, yırtılma sertliği ve yırtılma noktasındaki enerjinin etkisi Weka programında çizilen grafikler aracılığıyla tespit edilmeye çalışılmıştır (Şekil 2). Gerçek renk sınıfları ile algoritmaya göre belirlenen tahmin sınıflarının X-Y düzlemindeki dağılımı incelendiğinde kabuk renkleri bakımından domateslerin birbirinden tamamen homojen bir şekilde ayrıldığı görülmüştür (Şekil 2a). Kabuk yırtılma noktasındaki enerji değerlerine göre oluşan renk sınıflarının dağılımı ise Şekil 2b ’de verilmiştir. Dağılımdaki sınıf öbeğleşmesinden de anlaşılacağı üzere, kırılma ve yeşil renk dönemlerinde enerji değerlerinin diğer dönemlere göre daha yüksek olduğu belirlenmiştir (Şekil 2b). Aynı durum sertlik ve kuvvet değerleri içinde tespit edilmiştir (Şekil 2c, 2d). Mekanik özelliklerin değerlerinin domatesin en olgun olduğu kırmızı renkli sınıfa doğru azalış eğilimi göstermesi ise incelenen bu faktörlerin renk oluşumu üzerinde etkili olduğu ile açıklanabilir. Bu amaçla her bir özelliğin renk sınıfları üzerindeki etkisinin derecesi ve yönü domateslerin bireysel olarak Minolta ölçeğine göre belirlenen a\*/b\* değerlerinden yararlanılarak belirlenmiş ve Çizelge 8’de verilmiştir.

Domatesin renk olgunlukları ile kabuk yırtılma noktasındaki kuvvet (RF) değerleri arasındaki ilişki düzeylerinin r: (-0,672) - (0,31), sertlik (F) ile r: (-0,571) - (-0,113) ve kabuk yırtılma noktasındaki enerji (RE) ile r: (-0,541) - (0,194) değer aralığında değiştiği tespit edilmiştir (Çizelge 8). Ayrıca kabuk yırtılma

noktasındaki kuvvetin ve sertlik özelliklerinin yeşil, açık kırmızı ve renk dönüşüm dönemlerinde etkisinin önemsiz ( $P>0,05$ ), diğer dönemlerde ise önemli ( $P<0,01$ ) olduğu belirlenmiştir (Çizelge 8). Kabuk yırtılma noktasındaki enerji ise yalnızca pembe ve renk dönüşüm safhalarında etkili ( $P<0,01$ ) bulunmuş diğer renk oluşum safhalarında ise önemsiz olduğu tespit edilmiştir (Çizelge 8).

### Tartışma ve Sonuç

Araştırmada kullanılan toplam 352 adet domates renk uzayı ölçüm değerlerinin sınıflandırılmasında K-Star, Rastgele Orman ve Karar Ağacı (C4.5) algoritmaları kullanılmış ve en başarılı sınıflandırıcının belirlenmesinde modellerin hata varyans kriterlerinden yararlanılmıştır. Sınıflandırıcılar içerisinde başarı performansı diğerlerine nazaran daha yüksek bulunan algoritmanın örnek tabanlı K-Star Algoritması [MAE: 0,004; RMSE: 0,006; %RAE: 1,73 ve %RRSE: 1,70] olduğu tespit edilmiştir (Çizelge 6-7). Veri setine algoritmaların tek tek uygulanması sonucunda bulunan doğruluk oranları ise sırasıyla, %70,74; %98,30 ve %100 oranlarında bulunmuştur. Hız ve sınıflandırma modelinin doğruluk oranı bakımından yüksek bulunan K-Star algoritmasını ikinci olarak Rastgele Orman algoritması izlemiştir (Çizelge 6-7). Oluşan bu durum literatürde Rastgele Orman algoritması için diğer algoritmalarla göre daha hızlı olduğunu ve yüksek doğruluk oranı gösterdiğini vurgulayan araştırmacıların bildirimleri ile uyumaktadır (Watts ve Lawrence, 2008; Akar ve Güngör, 2010). Ayrıca bazı araştırmacılar, Şencan (2004)'in karar ağacı yöntemiyle tarımsal ürünleri sınıflandırmada %91,10 ve %66,15; Kavdir ve Guyer (2004)'in, Bayes ve K en yakın komşuluk ile elma çeşitlerini ayırtmada %84-89 arasında doğruluk oranlarına ulaştıklarını bildirmeleri veri madenciliğinin sınıflandırmada etkili olduğunun bir göstergesi olarak kabul edilebilir. Bu durum ise farklı

tarımsal ürünlerin sınıflandırılmasında, klasik yöntemlerin varsayımlarının yerine gelmediği durumlarda ve büyük çaplı veri setlerinin değerlendirilmesinde veri madenciliğinin alternatif bir yöntem oluşturabileceğini göstermektedir.

Araştırmada en iyi sınıflandırıcı olarak belirlenen örnek tabanlı algoritmalarından K-Star algoritmasına göre mekanik özelliklerin sınıflandırılması sonucunda ise domateslerin %16,19'nun yeşil, %6,81'nin renk kırılma, %26,42'sinin pembe, %19,03'nün açık kırmızı, %9,95'nin renk dönüşüm ve %21,60'nun da kırmızı renk safhalarında oldukları belirlenmiştir (Çizelge 5). Kabuk yırtılmadaki oluşan enerji düzeyi ise en düşük [0,002-0,01] kırmızı renkli domateslerde en yüksek [0,031- 0,90] ise renk kırılma dönemlerinde olduğu bulunmuştur (Şekil 2b). Söz konusu özellik için Sirisomboon ve ark. (2012)'nin bildirmiş olduğu absorbe edilen enerjinin en fazla kırmızı domateslerde olduğu bilgisi ile uyumsuz bulunmuştur. Bunun ise araştırmada kullanılan domates örneklerinin renk ölçüm değerlerinin ilgili renk safhalarında farklı miktarlarda (adet) bulunmasından kaynaklandığı düşünülmektedir. Domatesin sertlik değerleri ise en yüksek [2,05-5,41] yeşil renklilerde, en düşüğe kırmızı renklilerde [0,68-2,08] tespit edilmiş (Şekil 2c), domates olgunlaştıkça sertliğinin de azaldığı görülmüştür. Benzer sonuçlara ulaşan, Sirisomboon ve ark. (2012) tarafından da bu çalışmada olduğu gibi sertlik değerinin yeşilden kırmızıya doğru bir azalış gösterdiği tespit edilmiştir. Kabuk yırtılma noktasındaki kuvvet değerlerinin ise en yüksek [12,1-24,126] yeşil, en düşük [2,62-8,31] kırmızı renkli domateslerde olduğu bulunmuştur (Şekil 2d). Bu durum, Olorunda ve Tung (1985)'un çalışmalarında bildirdiği olgunluk derecesindeki artışın kuvvet, deformasyon ve sertlik değerlerinde belirgin bir azalmaya neden olduğu bilgileriyle tam uyumlu bulunmuştur.

Çizelge 5 Bağımlı ve bağımsız değişkenlere ait tanımlayıcı istatistikler

Mekanik özellikler	N	Ortalama	Standart hata	Max	Min
Kabuk yırtılma noktasındaki kuvvet (N) (RF)	352	8,91	5,67	24,126	2,62
Kabuk yırtılma sertliği (N mm <sup>-1</sup> ) (F)	352	1,968	0,955	5,41	0,68
Kabuk yırtılma noktasındaki enerji (N mm) (RE)	352	23	19	90	3

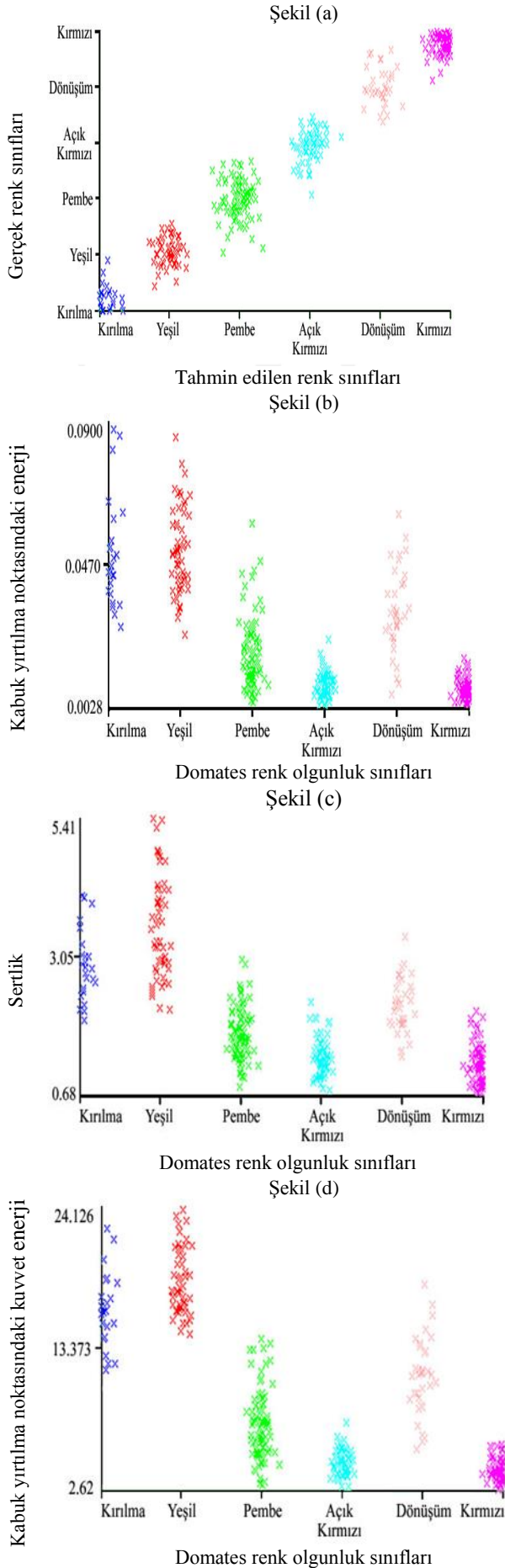
Renk sınıfları	Yeşil	Kırılma	Dönüşüm	Pembe	Açık kırmızı	Kırmızı
N	57	24	35	93	67	76
%	16,19	6,81	9,95	26,42	19,03	21,60

Çizelge 6 Sınıflandırıcı modellerin hata performans kriterlerinin değerleri

Sınıflandırıcılar	Kappa istatistiği	MAE	RMSE	% RAE	% RSSE	Hata oranı %	Doğruluk %	Hesaplama zamanı(sn)
Kstar	1	0,004	0,006	1,73	1,70	0	100	0,01
Karar ağacı (C4.5)	0,63	0,13	0,26	50,81	70,31	29,26	70,74	0,08
Rastgele Orman	0,97	0,05	0,12	20,38	34,58	1,70	98,30	0,03

Çizelge 7 Sınıflandırıcıların model başarımları ölçütleri

Sınıflandırıcılar	Duyarlılık	Keskinlik	F ölçütü
K-Star	1	1	1
Karar ağacı (C4.5)	0,707	0,726	0,703
Rastgele orman	0,983	0,984	0,983



Şekil 2 Domatesin renk olgunlukları üzerinde mekanik özelliklerin etkisinin K-Star algoritmasına göre dağılımları

Çizelge 8 Korelasyon değerleri

Renk sınıfları	RF	F	RE
Yeşil	0,31 <sup>ös</sup>	-0,151 <sup>ös</sup>	0,194 <sup>ös</sup>
Kırılma	-0,535 <sup>**</sup>	-0,541 <sup>**</sup>	-0,172 <sup>ös</sup>
Dönüşüm	-0,320 <sup>ös</sup>	-0,113 <sup>ös</sup>	-0,353 <sup>**</sup>
Pembe	-0,672 <sup>**</sup>	-0,571 <sup>**</sup>	-0,541 <sup>**</sup>
Açık kırmızı	-0,232 <sup>ös</sup>	-0,172 <sup>ös</sup>	-0,190 <sup>ös</sup>
Kırmızı	-0,310 <sup>**</sup>	-0,467 <sup>**</sup>	-0,122 <sup>ös</sup>

<sup>ös</sup>: P>0.05; <sup>\*\*</sup>: P<0.01, RF: Kabuk yırtılma kuvveti,

F: Kabuk yırtılma sertliği, RE: Kabuk yırtılma enerjisi

Araştırmada tespit edilen bir diğer önemli bulgu da, kabuk yırtılma noktasındaki kuvvetin en fazla kırmızı renklilerde ( $r: -0,31$ ), en az da pembe renklilerde ( $r: -0,0672$ ) etkili olduğunun ( $P < 0,01$ ) belirlenmesi olmuştur (Çizelge 8). Aynı durum sertlik değeri içinde tespit edilmiş olup, absorbe edilen enerji içinde en yüksek renk dönüşüm ( $r: -0,353$ ) safhasındaki domatesler üzerinde, en düşük ise pembe renklilerde ( $r: -0,541$ ) etkisinin önemli ( $P < 0,01$ ) olduğu belirlenmiştir (Çizelge 8). Bu durum ise domatesin renk olgunlukları ile mekanik özellikler arasında negatif bir ilişki olduğunu, bu değerlerin artmasına bağlı olarak renk olgunluklarının geciktiğini göstermiştir.

Bu çalışma ile domatesin renk olgunluk derecesi üzerinde bazı mekanik özelliklerin etkisi veri madenciliği yönteminin K-Star, Karar ağaçları (C4.5) ve Rastgele orman algoritmalarının kullanılmasına belirlenmiştir. Sonuç olarak domatesteki farklı renk olgunlukları üzerinde kabuk yırtılma kuvveti, sertlik ve kabuk yırtılma noktasındaki enerjinin etkili olduğu yeterli ve kaliteli domates üretiminde söz konusu özelliklere dikkat edilmesi gerektiği belirlenmiştir.

## Kaynaklar

- Aha DW, Kibler D, Albert MK. 1991. Instance-based learning algorithms. *Machine Learning* 6: 37–66.
- Akar Ö, Güngör O. 2010. Rastgele orman sınıflandırıcısı ile arazi kullanım alanlarının belirlenmesi, III. Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu, Geze, Kocaeli, 11-13 Ekim 2010, Bildiriler Kitabı, sf: 142-152.
- ASAE Standards 2009: [ASAE Standards (2009). Compression test of food materials of convex shape. ASAE S368.4 DEC 2000 (R2008). American Society of Agricultural and Biological Engineers, St. Joseph, Michigan, USA, pp. 678-686].
- Batu A. 2004. Determination of acceptable firmness and colour values of tomatoes. *Journal of Food Engineering* 61: 471-475.
- Biggs D, Ville BD, Suen E. 1991. A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics* 18: 49-62.
- Boyacıoğlu H, Güneri P. 2006. Sağlık Araştırmalarında Kullanılan Temel İstatistik Yöntemler, Hacettepe Dış Hekimliği Fakültesi Dergisi, 30: 33-39
- Braga GC, Couto SM, Hara T, Neto JTPA. 1999. Mechanical behaviour of macadamia nut under compression loading. *Journal of Agricultural Engineering Research*, 72: 239-245.
- Camdeviren HA, Yazıcı AC, Akkus Z, Bugdayci R, Sungur MA. 2007. Comparison of logistic regression model and classification tree: an application to postpartum depression data. *Expert Systems With Applications* 32: 987–994.
- Coşkun C, Baykal. A. 2011. Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması, Akademik Bilişim'11 - XIII. Akademik Bilişim Konferansı Bildirileri, Malatya, 2 - 4 Şubat 2011 İnönü Üniversitesi, S. 51-58.

- Çölkesen. 2010. Uzaktan algılamada ileri sınıflandırma tekniklerinin karşılaştırılması ve analizi, Yüksek Lisans Tezi, Gebze Yüksek Teknoloji Enstitüsü Mühendislik ve Fen Bilimleri Enstitüsü, Gebze, 153s.
- Er O, Cetişli B, Sofu MM, Kayacan MC. 2013. Gerçek zamanlı otomatik elma tasnifleme. Suleyman Demirel University Journal of Natural and Applied Science 17: 31-38.
- Eraltan FM. 2005. Şeftalinin mekanik özellikleri üzerine çeşit ve depolama süresi etkilerinin araştırılması, Yüksek Lisans tezi, Çukurova Üniv. Fen Bilimleri Ens., Adana,39s.
- Jay S, Lawrence R, Repasky K, Keith C. 2009. Invasive species mapping using low cost hyper spectral imagery, ASPRS 2009 Annual Conference, March 9-13, Baltimore, Maryland.
- Kara M, Turgut N. 1988. Erzurum yöresinde yetiştirilen patates çeşitlerinin önemli bazı mekanik özelliklerinin saptanması üzerine bir araştırma. Tarımsal Mekanizasyon 11. Ulusal Kongresi, Erzurum, (Tarih 10-12 Ekim 1988) 11. Ulusal Kongresi Bildiri Kitabı, sf: 302-313.
- Kavdir I, Guyer DE. 2004. Comparison of artificial neural networks and statistical classifiers in apple sorting using textural features. Biosystems Engineering 89: 331-344.
- Kavzoğlu T, Çölkesen İ. 2010. Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği, Harita Teknolojileri Elektronik Dergisi Cilt: 2, No: 1, 2010 (36-45).
- Keskin S. 2004. 2 x 2 Tablolarında Bazı Örnek Genişlikleri ve I. Tip Hata Seviyeleri (a) İçin Kappa (K) İstatistiğine Ait Ampirik Olarak Gerçekleşen Kritik Değerler, Tarım Bilimleri Dergisi, 10: 169-173.
- Lee KC, Han I, Kwon Y. 1996. Hybrid neural network models for bankruptcy predictions. Decision Support Systems 18: 63-73.
- Loh WY, Shih YS. 1997. Split selection methods for classification trees. Statistica Sinica 7: 815-840.
- Olorunda AO, Tung MA. 1985. Simulated transit studies on tomatoes; effects of compressive load, container, vibration and maturity on mechanical damage. Journal of Food Technology 20: 669-678.
- Sencan S. 2004. Decision Tree Classification of Multi-Temporal Images For Field-Based Crop Mapping, ODTU Fen Bilimleri Enstitüsü Yüksek Lisans Tezi. s. 139, Ankara.
- Sirisomboon P, Tanaka M, Kojima T. 2012. Evaluation of tomato textural mechanical properties. Journal of Food Engineering 111: 618-624.
- Vursavuş K, Özgüven F. 2007. Golden Delicious, Stark Krimson ve Granny Smith Elma Çeşitlerinin Çarpma Zedelenme Hassasiyetinin Belirlenmesi. Tarımsal Mekanizasyon 24. Ulusal Kongresi, Antakya 5-6 Eylül, 24. Ulusal Kongresi Bildiri Kitabı, sf:268-278.
- Watts JD, Lawrence RL. 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).
- Weka 3.4.5. 2013. <http://kent.dl.sourceforge.net/sourceforge/weka/weka-3-4-5jre.exe>. Erisim Tarihi: 25.01.2013.
- Yurtlu YB, Erdoğan D. 2005. Depolama süresinin bazı hıyar çeşitlerinde mekanik özelliklere olan etkisinin belirlenmesi. Tarım Bilimleri Dergisi 11: 251-256.