

## Renk Momentleri ve Destek Vektör Makineleri Kullanarak Haploid Mısır Tohumlarının Tanımlanmasında Renk Uzaylarının Sınıflandırma Performansına Etkisinin Karşılaştırılması

Yahya ALTUNTAŞ<sup>1\*</sup>, Adnan Fatih KOCAMAZ<sup>2</sup>

<sup>1,2</sup> Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, İnönü Üniversitesi, Malatya, Türkiye  
<sup>\*1</sup> yahyaaltuntas@gmail.com, <sup>2</sup> fatih.kocamaz@inonu.edu.tr

(Geliş/Received: 01/07/2019;

Kabul/Accepted: 22/08/2019)

**Öz:** Katlanmış haploid mısır ıslah tekniği gerek ıslah süresini kısaltması, gerekse %100 homozigot hatlar elde edilmesi açısından, geleneksel ıslah yöntemleri ile karşılaştırıldığında önemli avantajlar sağlar. Katlanmış haploid mısır ıslah yönteminde önemli işlerden biri melezleme sonucunda elde edilen haploid ve diploid tohumların ayrıştırılmasıdır. Günümüzde bu işlemin elle yapılıyor olması, emek ve zaman kaybının yanı sıra yüksek sınıflandırma hatalarına neden olmaktadır. Bu çalışmada, haploid mısır tohumlarının tanımlanması için bilgisayar görmesine dayalı bir yöntem önerilmiştir. 3000 mısır tohumundan oluşan veri seti üzerinde yürütülen çalışmada, k-ortalamar kümeleme yöntemi ile mısır tohum embriyoları bölütlenmiştir. RGB, HSV ve Lab renk uzaylarında her renk kanalı için ilk dört dereceden renk momentleri çıkarılmıştır. Böylece her renk uzayı için 12 öznitelik kullanılmıştır. Elde edilen öznitelikler destek vektör makinesi ile sınıflandırılmıştır. Daha genel sonuçlar elde edebilmek için 10-kat çapraz doğrulama yöntemi 20 kez tekrar edilmiştir. Deneysel sonuçlara göre en iyi performans Lab renk uzayında %87,03 doğruluk, %80,15 duyarlılık ve %93,00 özgünlük değerleriyle ölçülmüştür.

**Anahtar kelimeler:** mısır, haploid tanıma, renk uzayları, renk momentleri, destek vektör makineleri.

### Comparison of the Effect of Color Spaces on Classification Performance in Identification of Haploid Maize Seeds Using Color Moments and Support Vector Machines

**Abstract:** Doubled haploid maize breeding technique provides significant advantages over conventional breeding methods both in terms of shortening breeding time and obtaining 100% homozygous lines. One of the important works in doubled haploid maize breeding method is the separation of haploid and diploid seeds obtained as a result of hybridization. Today, this process is performed by hand, as well as loss of labor and time causes high classification errors. In this study, a computer vision based method for identification of haploid maize seeds was proposed. In the study conducted on the dataset of 3000 maize seeds, seed embryos were segmented by k-means clustering method. In the RGB, HSV and Lab color spaces, the first four degrees of color moments are extracted for each color channel. Thus, 12 features were used for each color space. Obtained features were classified with support vector machine. To obtain more general results, the 10-fold cross-validation method was repeated 20 times. According to the experimental results, the best performance was measured with 87,03% accuracy, 80,15% sensitivity and 93,00% specificity in Lab color space.

**Key words:** maize, haploid identification, color spaces, color moments, support vector machines.

#### 1. Giriş

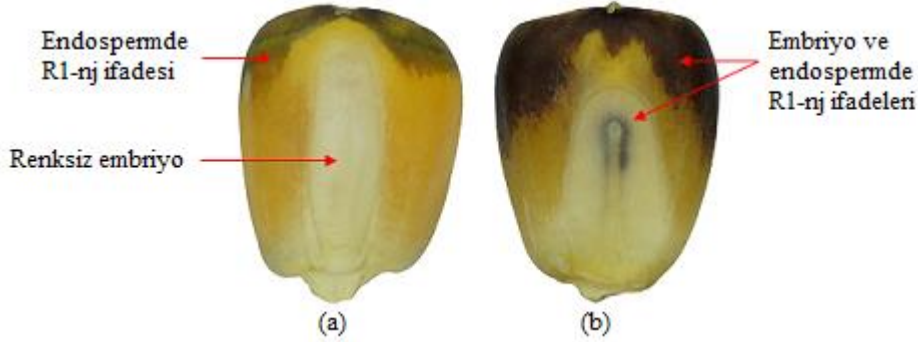
Diğer tüm kültür bitkilerinde olduğu gibi, artan dünya nüfusunun ihtiyaçlarını karşılayacak ve küresel iklim değişikliklerine uyum sağlayacak yüksek verimli, biyotik ve abiyotik stres koşullarına dayanıklı yeni mısır (*Zea mays* L.) çeşitlerinin geliştirilmesi bir zorunluluktur [1]. Katlanmış haploid (KH) mısır ıslah tekniği gerek ıslah süresini kısaltması, gerekse %100 homozigot hatlar elde edilmesi açısından, geleneksel ıslah yöntemleri ile karşılaştırıldığında önemli avantajlar sağlar. Geleneksel ıslah yönteminde, saf hatlar sürekli kendiyile melezlenerek geliştirilir. Bu işlem 6-7 yıl sürer ve işlem sonunda %99 homozigotluk düzeyinde hatlar elde edilir. Bununla birlikte, KH tekniği ile %100 homozigotluk düzeyinde hatların geliştirilmesi 1-2 yılda başarılabilir [2].

KH, tek n kromozoma sahip haploid embriyoların suni kromozom katlanmasıyla üretilen %100 homozigotluk düzeyinde hatlardır [2]. Bu işlem için öncelikle haploid bitkilerin elde edilmesi gerekir. Haploidler doğada %0,1 gibi çok düşük bir frekansta bulunmaktadır [3], bu nedenle pratik kullanıma uygun değildir [4]. Literatürde daha yüksek oranlarda haploid bitki elde etmek için yöntemler olmakla birlikte, in vivo maternal haploid yöntemi standart yöntem haline gelmiştir [4]. In vivo maternal haploid tekniğinde, daha yüksek oranlarda haploid bitkiler elde etmek için tozlayıcı olarak indirgeyici hat adı verilen özel genotipler kullanılmaktadır [5]. Mevcut indirgeyici

\* Sorumlu yazar: [yahyaaltuntas@gmail.com](mailto:yahyaaltuntas@gmail.com). Yazarların ORCID Numarası: <sup>1</sup> 0000-0002-7472-8251, <sup>2</sup> 0000-0002-7729-8322

hatlar sayesinde haploid bitki elde etme oranı %8 ve daha yüksek olduğu bildirilmiştir [6]. In vivo maternal haploid tekniği dört aşamadan oluşur: (i) haploidlerin elde edilmesi, (ii) haploidlerin diploidlerden ayrılması, (iii) suni kromozom katlanması ve (iv) nihai KH hattının üretilmesi için kendisiyle tozlaşması.

Suni kromozom katlanmasında haploidler kullanıldığından, kromozom katlanmasından önce küçük bir orana sahip olan haploidlerin diploidlerden ayrılması gerekir. Tohum döneminde haploidlerin diploidlerden ayrılması; suni kromozom katlanmasına izin vermek, maliyeti düşürmek, işçiliği azaltmak, sera çalışmalarını ve tarla alanını azaltmak gibi avantajlar sağlar [2]. Bu gerçek, hızlı ve kesin olarak haploid tohum seçiminin önemini vurgulamaktadır [7]. Haploidlerin diploidlerden ayrımını yapabilmek için farklı yağ içeriği seviyeleri [6], embriyo ağırlıkları arasındaki fark [8] ve kırmızı kök [9] gibi alternatif seçilebilir genetik markörler önerilmişse de, en yaygın olarak ve başarıyla kullanılan R1-Navajo (R1-nj) [10] renk markörüdür. Baskın R1-nj geni embriyo ve endosperimde kırmızı-mor renklenmeye sebep olur. Diploid tohumlar hem kaynak genotipin hem de indirgeyici hattın kromozomlarını içerdiğinden embriyo ve endospermlerinde R1-nj renklenmesi gözlenir. Haploid tohumlar ise, endospermlerinden kaynak genotip ve indirgeyici hattın kromozomlarını taşıırken, embriyolarında sadece kaynak genotipin kromozomlarını taşıdıklarından endospermlerinde R1-nj renklenmesi gözlenirken, embriyolarında bu renklenme meydana gelmez. Bu durum, haploid ve diploidlerin görsel olarak ayrılmasına olanak tanır. Haploid ve diploid mısır tohumlarının R1-nj renk markörüne göre görsel ayrımı Şekil 1'de gösterilmiştir [1].



Şekil 1. R1-nj renk markörüne göre (a) haploid ve (b) diploid mısır tohumlarının görsel ayrımı [1]

Haploid mısır tohumlarının diploidlerden ayrılması günümüzde elle yapılmaktadır. Emek yoğun ve zaman alan bu işlem, R1-nj renklenmesinin küçük bir noktadan tüm endospermi kaplayan renkliliğe kadar çeşitlilik göstermesi ve renklenmenin açıktan çok koyuya kadar çeşitlilik göstermesi nedeniyle yüksek hataya sebep olmaktadır. Bu nedenle, haploid tohumların tanımlanmasını otomatik olarak gerçekleştirecek bir sisteme ihtiyaç vardır [11]

Son birkaç yılda R1-nj renk markör sistemine göre haploid mısır tohumlarının otomatik seçimini başarmak için makine görmesine dayalı çalışmalar yayınlanmıştır. Boote vd. [12] floresan bazlı yeni bir yöntem önerdiler. Floresan görüntülemenin tohum işleme için mevcut teknolojilerle birleştirildiğinde yüksek sınıflandırma etkinliğine sahip bir yöntem olduğunu ifade etmişlerdir. Lin vd. [13] haploid mısır tohumlarının yakın kızılötesi spektrum özellikleri kullanılarak otomatik olarak seçilebileceğini bildirmişlerdir. Fuente vd. [14] VideometerLab 3 yazılımının performansını test etmişler ve %50'nin üzerinde doğrulukla mümkün olduğunu bildirmişlerdir. Altuntaş vd. [15] 5 öznitelik ve destek vektör makinesi kullandıkları yöntemle %81,36 doğruluk elde etmişlerdir. Altuntaş vd. [16] aynı veriseti üzerinde gri seviye eş oluşum matrisinden elde ettikleri doku özniteliklerini farklı makine öğrenmesi algoritmaları kullanarak sınıflandırmışlar ve %84,48 doğruluk elde ettiklerini bildirmişlerdir. Altuntaş vd. [1] evrimsel sinir ağları kullanarak 3.000 örnek üzerinde yürüttükleri çalışmada, haploid tanımlama başarımının %94,22 doğruluk oranıyla mümkün olduğunu bildirmişlerdir.

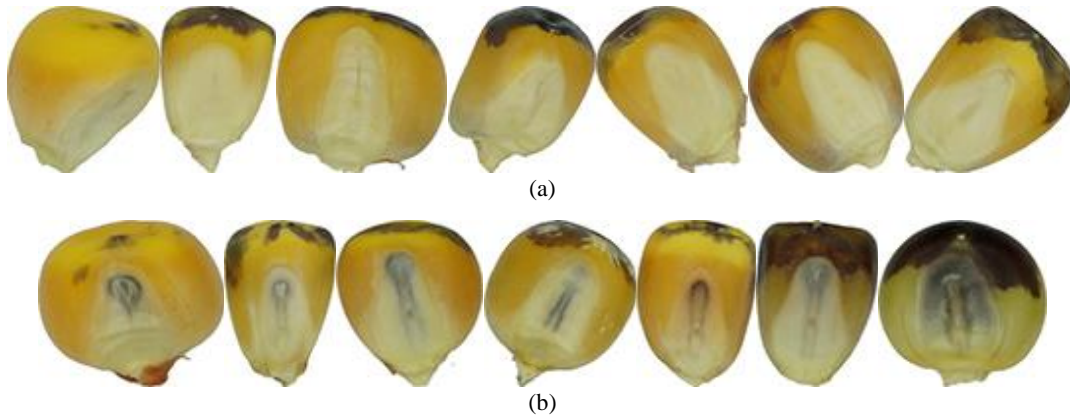
Bu çalışmada haploid ve diploid mısır tohumlarının R1-nj renk markör sistemine göre otomatik olarak tanımlanması için embriyo bölgesinden elde edilen ilk dört dereceden renk momentleri kullanılmış ve elde edilen öznitelik vektörü destek vektör makinesinin (SVM) eğitimi için kullanılmıştır. 3 farklı renk uzayından elde edilen öznitelikler karşılaştırılmış ve en uygun renk uzayının %87,03 doğruluk oranıyla Lab olduğu belirlenmiştir.

Bu makalenin yapısı şu şekilde düzenlenmiştir: 2. Bölümde kullanılan veri seti, embriyo bölütleme yöntemi, öznelik çıkarımı ve destek vektör makineleri anlatılmıştır. Elde edilen deneysel sonuçlar Bölüm 3’de verilmiştir. Bölüm 4’de önerilen modelin performansı değerlendirilmiş ve gelecek çalışmalara değinilmiştir.

## 2. Materyal ve Yöntem

### 2.1. Veri seti

Bu çalışmada, 3.000 mısır tohumundan oluşan açık erişimli “Haploid ve Diploid Mısır Tohumları Veri Seti” [1] kullanılmıştır. Veri seti 1.230 haploid ve 1.770 diploid mısır tohum görüntüsünden oluşmaktadır. Veri setini oluşturan mısır tohumları 107 kaynak genotip ile “RWS/RWK-76” [5] indirgeyici hatları arasında yapılan melezleme sonucunda elde edilmiştir. Haploid ve diploid mısır tohumlarının etiketlenmesinde R1-nj ifadeleri dikkate alınmıştır. Veri setini oluşturan mısır tohum görüntüleri, tohum boyutuna göre değişmekle birlikte 300x289 piksel ile 610x637 piksel aralığında değişmektedir. Veri setinden örnek haploid ve diploid mısır tohum görüntüleri Şekil 2’de gösterilmiştir.



Şekil 2. Veri setini oluşturan (a) haploid ve (b) diploid mısır tohumlarına ait örnek görüntüler

### 2.2. Embriyo Bölütleme

Daha önce de belirtildiği gibi, hem haploid hem de diploid mısır tohumlarının endospermelerinde baba ebeveynin kromozomlarını taşıdıkları için R1-nj renklenmesi gözlenmektedir. Bununla birlikte diploid tohumların embriyolarında R1-nj renklenmesi oluşurken, baba ebeveynin kromozomlarını taşımayan haploid tohumların embriyolarında R1-nj renklenmesi oluşmamaktadır. Bu sebeple sadece embriyo renklenmelerine bakılarak, mısır tohumları haploid ve diploid olarak ayrılabilir.

Öznelikler çıkarılmadan önce veri setini oluşturan örneklerin embriyoları bölütlenmiştir. Embriyo bölütleme işlemi için k-ortalamlar kümeleme yöntemi kullanılmıştır. Algoritma sonunda mısır tohumu embriyo ve endosperm olmak üzere iki parçaya ayrılmıştır. Embriyo bölütleme işlemine başlanmadan önce görüntünün arkaplanı temizlenmiş ve sadece tohum pikselleri hesaplamalarda kullanılmıştır. Denklem 1’e göre görüntünün B kanalı 150 eşik değeri ile eşiklenerek arka plan görüntüden çıkarılmıştır.

$$piksel_{i,j} = \begin{cases} mısır\ tohumu, & B_{i,j} \leq 150 \\ arkaplan, & deęilse \end{cases} \quad (1)$$

Burada  $piksel_{i,j}$  ikili görüntüde  $i$ . satır ve  $j$ . sütun piksel değerini,  $B_{i,j}$  orijinal görüntünün  $i$ . satır ve  $j$ . sütundaki pikselin B değerini temsil etmektedir.

k-ortalamlar kümeleme yöntemi bir başlangıç çözümüne ihtiyaç duyar. R1-nj renklenmesi göz ardı edilirse, mısır tohumu embriyosu krem renkli ve endosperm sarı renklidir. Bu bilgi kullanılarak RGB renk kanalları analiz

edilmiş ve ampirik olarak B kanalı bölütleme işlemi için en uygun kanal ve eşik değeri 100 olarak ayarlanmıştır. Denklem 2 kullanılarak başlangıç çözümü bulunmuştur.

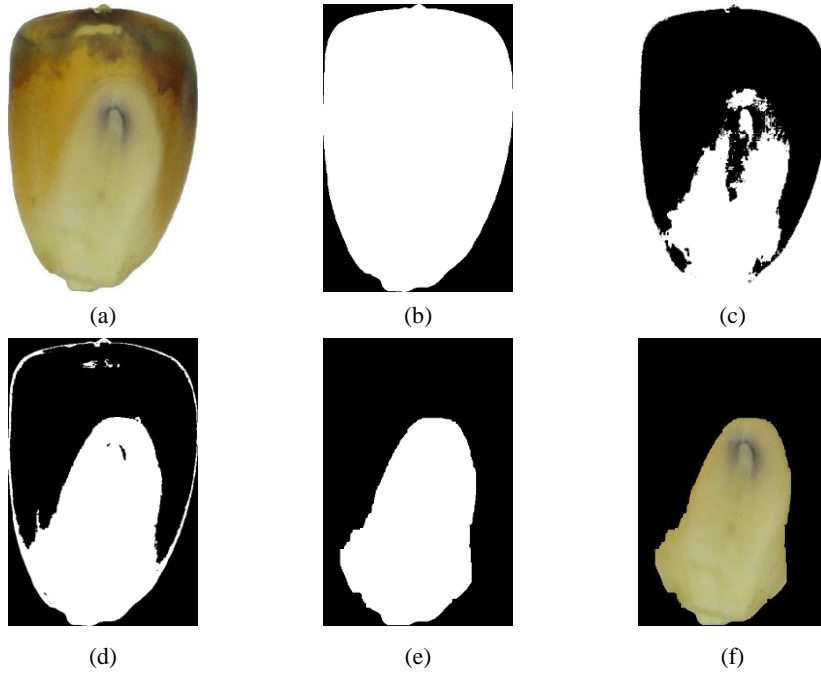
$$piksel_{i,j} = \begin{cases} embriyo, & B_{i,j} > 100 \\ endosperm, & \text{değilse} \end{cases} \quad (2)$$

Yapılan eşikleme neticesinde embriyo ve endosperm olarak seçilen piksellerin ortalama B değerleri küme merkezleri olarak hesaplanmıştır. Daha sonra tekrarlamalı olarak her bir pikselin B değerinin küme merkezlerine olan öklid uzaklıkları hesaplanmış ve pikseller en yakın oldukları kümeye atanmıştır. Denklem 3 kullanılarak piksel atamaları gerçekleştirilmiştir.

$$piksel_{i,j} = \begin{cases} embriyo, & d_E[(B_{i,j})(B_{embriyo})] \leq d_E[(B_{i,j})(B_{endosperm})] \\ endosperm, & \text{değilse} \end{cases} \quad (3)$$

Burada  $d_E$   $i$ . satır ve  $j$ . sütunda bulunan pikselin B değerinin küme merkezlerine olan öklid uzaklığını temsil etmektedir. Piksel atama işleminin ardından küme merkezleri yeniden hesaplanmış ve piksel atama işlemi tekrarlanmıştır. Tekrarlama işlemi küme merkezleri değişmeyinceye kadar devam ettirilir. 5 tekrar nihai çözüm için yeterli olduğundan, bu çalışmada, tekrar sayısı 5 ile sınırlandırılmıştır.

Elde edilen çözüm, tohum kenarlarındaki parlamadan kaynaklanan gürültüden temizlemek için aşındırma ve genişletme morfolojik işlemlerine tabi tutulmuştur. Embriyo bölütleme işlem adımları Şekil 3’de gösterilmiştir.



**Şekil 3.** Embriyo bölütleme işlem adımları: (a) orijinal görüntü, (b) arkaplan temizlenmiş ikili görüntü, (c) başlangıç çözümü, (d) kümeleme algoritması sonucunda elde edilen ikili görüntü, (e) aşındırma ve genişletme morfolojik işlemleri uygulanmış ikili görüntü, (f) nihai çözüm

### 2.3. Renk Uzayları

Renk uzayı, renk bilgisini üç veya dört farklı renk bileşeni olarak temsil eden matematiksel bir modeldir [17]. Aslında bir renk uzayı, her rengin tek bir nokta ile temsil edildiği bir koordinat sistemi içerisindeki bir alt uzayın

tanımlanmasıdır [18]. Bir renk uzayının amacı, bir takım standartlar ile renk tanımlamalarını kolaylaştırmaktır [18]. Farklı uygulamalar için kullanılacak farklı renk uzayları tanımlanmıştır. Bu çalışmada RGB, HSV ve Lab renk uzayları kullanılmıştır.

### 2.3.1. RGB Renk Uzayı

RGB renk uzayı yaygın olarak kullanılmaktadır ve dijital görüntüleri depolamak ve temsil etmek için varsayılan renk uzayıdır [17]. RGB renk uzayı üç farklı renk kanalına sahiptir: kırmızı, yeşil ve mavi. Bu üç temel renk karıştırılarak herhangi bir renk elde edilebilir. Bu üçlü 0 ile 255 arasında değişen bir renk değeri alır. RGB renk uzayında  $256^3 = 16.777.216$  farklı renk temsil edilebilir. RGB renk uzayı, renk bilgisinin gerektiğinde farklı renk uzaylarına dönüştürülebildiği temel renk uzayıdır [19].

### 2.3.2. HSV Renk Uzayı

HSV renk uzayı, bilgisayar grafiklerinde sıkça kullanılır ve renkleri tanımlamanın sezgisel bir yoludur [19]. HSV renk uzayı renk özü, doygunluk ve değer olmak üzere üç kanaldan oluşur. Renk özü, baskın dalga uzunluğunu tanımlar ve 0 ile 360 arasında değişen açısal bir değerdir. Doymunluk, görüntüdeki gri seviye değerini tanımlar ve genel olarak 0 ile 100 arasında değişir. Değer ise yoğunluğu (parlaklığı) tanımlar ve %0 ile %100 arasında değişir, burada %0 en karanlık ve %100 en parlak renge karşılık gelir.

### 2.3.3. Lab Renk Uzayı

Lab renk uzayının en önemli özelliği, renkleri cihazdan bağımsız olarak tanımlanmasıdır. Bu özelliği, farklı renklerin farklı cihazlar arasında iletişim imkanı sağlaması anlamına gelmektedir [20]. Lab kullanılarak, spektrumdaki tüm renkler tanımlanabilir. L aydınlığı tanımlar ve 0 ile 100 arasında değerler alır. a ve b renk boyutlarını ifade eder ve her eksen için pozitif ile negatif arasında değerler alır. Pozitif a değerleri kırmızı miktarını belirtirken, negatif değerler yeşil miktarını belirtir. Pozitif b değerleri sarı miktarını belirtirken, negatif b değerleri mavi miktarını belirtir [20].

## 2.4. Öznitelik Çıkarımı

Renk momentleri, renkli bir görüntünün renk dağılımının karakteristiğini temsil eden özniteliklerdir. Her renk uzayından her renk kanalı için, ilk dört dereceden renk momentleri (ortalama, standart sapma, çarpıklık ve kurtosis) öznitelik olarak kullanılmıştır. Böylece, veri setini oluşturan her bir örnek, her renk uzayı için toplam 12 öznitelik kullanılarak temsil edilmiştir.

İlk dört dereceden renk momentleri şu şekilde tanımlanır:

$$\mu = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n piksel_{ij} \quad (4)$$

$$\sigma = \sqrt{\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (piksel_{ij} - \mu)^2} \quad (5)$$

$$s = \sqrt[3]{\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (piksel_{ij} - \mu)^3} \quad (6)$$

$$k = \sqrt[4]{\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (piksel_{ij} - \mu)^4} \quad (7)$$

Burada,  $\mu$  ortalamayı,  $\sigma$  standart sapmayı,  $s$  çarpıklığı ve  $k$  kurtosisi ifade etmektedir.  $m \times n$  çözünürlükte bir görüntüde  $piksel_{ij}$  ilgili renk kanalında  $i$ . satır ve  $j$ . sütundaki piksel değerini temsil etmektedir.

## 2.5. Sınıflandırma

Bu çalışmada, elde edilen özneliklerin sınıflandırılmasında SVM kullanılmıştır. SVM, pek çok gerçek dünya probleminin çözümünde üstün sınıflandırma performansı göstermektedir [21]. Ayrıca, tarımsal verilerin sınıflandırılmasında, diğer geleneksel makine öğrenmesi yöntemlerine göre daha üstün sonuçlar verdiği bildirilmiştir [22].

SVM, gözetimli bir sınıflandırma yöntemidir. SVM, bir sınıfın tüm veri noktalarını diğer sınıflardan ayıran en iyi hiper düzlemi bularak verileri sınıflandırır [23]. SVM için en iyi hiper düzlem, iki sınıf arasında en büyük aralığa sahip olan doğrusal ayırıcı olarak ele alınmaktadır. Aralık, iç veri noktaları bulunmayan hiper düzleme paralel doğrusal ayırıcının maksimum genişliğini ifade eder.

Eğitim verisi,  $y_j$  kategorileriyle birlikte bir dizi nokta (vektör)  $x_j'$  dir. Bazı boyutlar için  $d$ ,  $x_j \in R^d$  ve  $y_j = \pm 1$ ' dir. Bir hiper düzlemin denklemi aşağıdaki gibidir.

$$f(x) = x'\beta + b = 0 \quad (8)$$

Bu denklemde  $\beta \in R^d$  ve  $b$  gerçek bir sayıdır. Aşağıdaki problem, en iyi ayırıcı hiper düzlemini (yani karar sınırını) tanımlar. Bu problem kapsamında tüm veri noktaları  $(x_j, y_j)$  için,  $\|\beta\|$  parametresini en aza indiren  $\beta$  ve  $b$  işaretlerini bulmak amaçlanır.

$$y_j f(x_j) \geq 1 \quad (9)$$

Destek vektörleri sınırdaki  $x_j'$  değerleridir, bu yüzden  $y_j f(x_j) = 1$ ' dir. Matematiksel uygunluk için, problem genellikle  $\|\beta\|$  seviyesinin minimize edilmesi eşdeğer problemi olarak verilir. Bu ikinci dereceden bir programlama problemine karşılık gelir. Optimal çözüm  $(\hat{\beta}, \hat{b})$  bir vektör  $z'$  nin aşağıdaki gibi sınıflandırılmasını sağlar:

$$class(z) = sign(z'\hat{\beta} + \hat{b}) = sign(\hat{f}(z)) \quad (10)$$

$f(z)$  sınıflandırma puanıdır ve karar sınırından olan  $z$  mesafesini temsil eder. Bu çalışma kapsamında bu mesafe değerini en iyi şekilde bulmaya çalışan destek vektör makinesi yöntemi mısır tohumlarını haploid ve diploid olarak ayıran sınıflandırıcı olarak kullanılmıştır.

### 3. Bulgular

Bu çalışmada, haploid ve diploid mısır tohumlarının sınıflandırılması için, bölütleme, öznelik çıkarımı ve sınıflandırma olmak üzere üç ana aşamadan oluşan bir model önerdik. Bölütleme, öznelik çıkarımı ve sınıflandırma işlemlerinin tamamı MATLAB © (2016a) ortamında gerçekleştirildi.

Daha öncede belirtildiği gibi, hem haploid hem de diploid mısır tohumlarının endospermelerinde R1-nj renklenmesi oluştuğundan, endospermdeki renklenme ayırt edici değildir. Ancak, diploid mısır tohumlarının embriyolarında R1-nj renklenmesi meydana gelirken, haploid mısır tohumlarının embriyolarında bu renklenme oluşmamaktadır. Bu durum, haploid ve diploid mısır tohumlarının görsel olarak sınıflandırılmasına olanak sağlamaktadır. Bu sebeple, öznelik çıkarma aşamasından önce mısır tohum embriyoları k-ortalama kümeleme yöntemiyle bölütlendi. Bölütleme işlemi için görüntünün B kanalı kullanıldı.

RGB renk uzayında bölütlenen embriyo görüntüleri HSV ve Lab renk uzaylarına dönüştürüldü. Üç renk uzayının üçer renk kanalından ilk dört dereceden renk momentleri hesaplandı. Böylece, veri setindeki her örnek her renk uzayı için sadece 12 renk momenti kullanılarak temsil edildi.

Sınıflandırma aşamasında, veri setinin yapısı bize haploid ve diploid mısır tohumlarını doğrusal olarak ayırma imkanı vermediğinden, gaussian radyal tabanlı fonksiyon (RBF) çekirdeği olan SVM, haploid örnekleri tanımak için kullanılmıştır. RBF gibi çekirdek fonksiyonları, çalışılan verilerin, verilerin ayrılacağı yeni bir hiper düzlemine dönüştürülmesini sağlar. SVM için bir diğer önemli parametre, yanlış sınıflandırılmış verinin oranını kontrol etmek için kullanılan ceza faktörüdür (kutu kısıtı olarak da bilinir). Ağ yapılandırma aşamasında, SVM için sigma ve ceza parametresi çiftlerini sırasıyla 1 - 10 ile 0,1 - 0,9 aralıklarında araştırdık. Gerçekleştirilen ampirik deneyler sonucunda sigma ve ceza faktörü sırasıyla 3 ve 0,8 olarak ayarlandı.

Modellerin başarısını ölçmek için kullanılan performans ölçütleri, Şekil 4'de [24] verilen karmaşıklık matrisinden elde edilir. Karmaşıklık matrisi, doğru pozitif (TP), doğru negatif (TN), yanlış pozitif (FP) ve yanlış negatif (FN) olarak adlandırılan dört endeksten oluşur. Bizim çalışmamızda, doğru bir şekilde tanımlanmak istenen haploidler pozitif sınıf, diploidler ise, negatif sınıf olarak ele alındı. Karmaşıklık matrisinde TP ve TN, doğru şekilde tanımlanmış haploid ve diploid örneklerinin sayısına karşılık gelirken, FP ve FN, yanlış sınıflandırılmış haploid ve diploid örneklerinin sayısını temsil etmektedir.  $\pm R$ , veri setindeki pozitif (haploid) ve negatif (diploid) sınıflarına ait örnek sayılarını temsil ederken,  $\pm P$  önerilen model tarafından tahmin edilen pozitif (haploid) ve negatif (diploid) sınıflarına ait örnek sayılarını, ifade etmektedir.

		+R	-R	
+P	TP	FP	TP+FP	
-P	FN	TN	FN+TN	
	TP+FN	FP+TN	N	

Şekil 4. Karmaşıklık matrisi [24]

Renk uzaylarının karşılaştırılması için kullanılan performans ölçütleri, farklı renk uzaylarından elde edilen farklı öznelik vektörlerine dayanan modelleri değerlendirmek için kullanılmıştır. Doğruluk (Acc), modelin ortalama performansını verir. Duyarlılık (Se), haploid örnekleri tanıma model performansını gösterirken, Özgünlük (Sp) diploid örnekler üzerindeki model performansını gösterir. Örneklerin sınıflar arasındaki dağılımı eşit değildir. Böyle bir durumda, metriklerin yorumlanması bir hataya yol açabilir ve bu nedenle kalite endeksinin (QI) ve F-ölçütünün dikkate alınması model performanslarının daha doğru bir şekilde karşılaştırılmasına imkan verir. QI, Se ve Sp'in geometrik ortalamasını ifade eder. F-ölçütü, model performansının karakterizasyonunu değerlendirmek için yararlı bir ölçümdür. Çalışmada kullanılan performans metrikleri aşağıda verilmiştir [25, 26]:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \quad (11)$$

$$Se = \frac{TP}{TP + FN} \quad (12)$$

$$Sp = \frac{TN}{TN + FP} \quad (13)$$

$$QI = \sqrt{Se * Sp} \quad (14)$$

$$F - \text{ölçütü} = \frac{2TP}{2TP + FP + FN} \quad (15)$$

Sınıflandırıcı eğitim ve test aşamalarında, daha genel sonuçlar elde etmek için 10-kat çapraz doğrulama tekniği kullanıldı. Ayrıca, tüm deneysel kurulum 20 kez tekrarlandı. Bu işlem sonucunda toplam 60 test işlemi gerçekleştirildi. Deneylere göre, elde edilen karmaşıklık matrisleri Tablo 1'de sunulmuştur. Ek olarak, performans metriklerinin ortalama ve standart sapma değerleri Tablo 2'de rapor edilmiştir.

**Tablo 1.** Renk uzaylarının karmaşıklık matrisleri

	TP	FP	TN	FN
RGB	22.084	2.516	29.245	6.155
HSV	22.235	2.365	29.056	6.344
Lab	22.353	2.247	29.864	5.536

**Tablo 2.** Performans metrikleri

	Acc (%)	Se (%)	Sp (%)	QI (%)	F-ölçütü (%)
RGB	85,55±0,20	78,20±0,24	92,08±0,21	84,86±0,20	83,59±0,23
HSV	85,49±0,18	77,80±0,24	92,47±0,21	84,82±0,19	83,62±0,21
Lab	87,03±0,13	80,15±0,23	93,00±0,14	86,34±0,13	85,17±0,13

Ortalama ± Standart sapma

DeneySEL sonuçlara göre, en iyi sınıflandırma performansı Lab renk uzayından çıkartılan öz nitelikler kullanılarak elde edildi. Acc, Se ve Sp değerleri sırasıyla %87,03, %80,15 ve %93,00 olarak gerçekleşmişti. RGB ve HSV renk uzaylarında elde edilen sınıflandırma sonuçları birbirlerine oldukça yakın olmakla birlikte, Lab renk uzayı ile karşılaştırıldığında nispeten düşüktü. RGB renk uzayında Acc, Se ve Sp değerleri sırasıyla %85,55, %78,20 ve %92,08 olarak ölçülmüştür. HSV renk uzayında Acc, Se ve Sp değerleri sırasıyla %85,49, %77,80 ve %92,47 olarak ölçülmüştür. Standart sapma değerleri tüm performans metrikleri için 0,13-0,24 aralığında değişmektedir. Bu düşük standart sapma değerleri önerilen modellerin stabil olduğunu göstermektedir.



#### 4. Sonuç

Haploid mısır tohumlarının tanınması, KH mısır ıslah yönteminin hızlandırılması ve etkinliğinin artırılması açısından oldukça önemli bir konudur. Günümüzde in vivo maternal haploid tekniği ile elde edilen mısır tohumlarının haploid, diploid ayrımı el ile yapılmaktadır. Bu seçimin el ile yapılması emek yoğun, zaman alan bir iş olmasının yanı sıra yüksek hata oranına sebep olmaktadır. Haploid tohumların manuel seçimi çok uzun zaman gerektirir ve bu işlem aynı zamanda yüksek seviyede bir hata içerir. Bu çalışmada, haploid mısır tohumlarını otomatik olarak tanımak için renk momenti özneliklerine dayalı yeni bir bilgisayar destekli yöntem önerilmiş ve 3 farklı renk uzayı karşılaştırmalı olarak test edilmiştir. Renk momenti öznelikleri ve SVM sınıflandırıcısı kullanılarak tatmin edici bir başarı sağlanmıştır. 3.000 örnekten oluşan veri seti üzerinde yürütülen çalışmada, elde edilen deney sonuçları, önerilen yaklaşımın haploid mısır tohumlarının tanınması için geleneksel yaklaşımdan daha pratik bir sistem sağlayabileceğini kanıtlamıştır. Deneyler bağlamında, RGB, HSV ve Lab renk uzaylarından elde edilen renk momenti öznelikleri incelenmiştir. Deney sonuçları, renk uzayları birbirleriyle karşılaştırıldıklarında en iyi sonuçların sırasıyla Lab, RGB ve HSV uzaylarından elde edildiğini göstermektedir. Sonuç olarak, önerilen yöntem haploid mısır tohumlarını %87,03 doğruluk, %80,15 duyarlılık ve %93,00 özgünlük ile ayırt edebilmektedir. Gelecekteki çalışmalarda, elde edilen deneysel sonuçların sahaya aktarılması için bir makine tasarımı gerçekleştirilecektir.

#### Kaynaklar

- [1] Altuntaş Y, Cömert Z, Kocamaz AF. Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach. *Comput Electron Agric* 2019; 163:104874.
- [2] Prasanna BM, Chaikam V, Mahuku G. *Doubled Haploid Technology in Maize Breeding: Theory and Practice*. Mexico: CIMMYT, 2012.
- [3] Geiger HH. *Doubled Haploids, Handbook of Maize*. New York, USA: Springer New York, 2009; 641-657.
- [4] Chidzanga C, Muzawazi F, Midzi J, Hove T. Production and use of haploids and doubled haploid in maize breeding: A review. *African Journal of Plant Breeding* 2017; 4: 201–213.
- [5] Röber FK, Gordillo GA, Geiger HH. In vivo haploid induction in maize - Performance of new inducers and significance of doubled haploid lines in hybrid breeding. *Maydica* 2005; 50: 275–283.
- [6] Melchinger AE, Schipprack W, Friedrich Utz H, Mirdita V. In vivo haploid induction in maize: identification of haploid seeds by their oil content. *Crop Sci* 2014; 54(4): 1497-1504.
- [7] Geiger HH, Gordillo GA, Koch S. Genetic Correlations among Haploids, Doubled Haploids, and Testcrosses in Maize. *Crop Sci* 2013; 53: 2313–2320.
- [8] Smelser A, Blanco M, Lübberstedt T, Schechert A, Vanous A, Gardner C. Weighing in on a method to discriminate maize haploid from hybrid seed. *Plant Breed* 2015; 134(3): 283–285.
- [9] Chaikam V, Martinez L, Melchinger AE, Schipprack W, Boddupalli PM. Development and validation of red root marker-based haploid inducers in maize. *Crop Sci* 2016; 56(4): 1678–1688.
- [10] Nanda DK, Chase SS. An Embryo Marker for Detecting Monoploids Of Maize (*Zea Mays L.*)1. *Crop Sci* 1966; 6(2): 213-215.
- [11] Melchinger AE, Schipprack W, Würschum T, Chen S, Technow F. Rapid and accurate identification of in vivo-induced haploid seeds based on oil content in maize. *Sci Rep* 2013; 3: 2129.
- [12] Boote BW, Freppon DJ, De La Fuente GN, Lübberstedt T, Nikolau BJ, Smith EA. Haploid differentiation in maize kernels based on fluorescence imaging. *Plant Breed* 2016; 135(4): 439–445.
- [13] Lin J, Yu L, Li W, Qin H, Method for Identifying Maize Haploid Seeds by Applying Diffuse Transmission Near-Infrared Spectroscopy. *Appl. Spectrosc* 2018; 72(4): 611–617.
- [14] De La Fuente GN, Carstensen JM, Edberg MA, Lübberstedt T. Discrimination of haploid and diploid maize kernels via multispectral imaging. *Plant Breed* 2017; 136(1): 50–60.
- [15] Altuntaş Y, Kocamaz AF, Cengiz R, Esmeray M. Classification of haploid and diploid maize seeds by using image processing techniques and support vector machines. In: 2018 26th Signal Processing and Communications Applications Conference (SIU); 2-5 May 2018; İzmir, Turkey. 1–4.
- [16] Altuntaş Y, Kocamaz AF, Cömert Z, Cengiz R, Esmeray M. Identification of haploid maize seeds using gray level co-occurrence matrix and machine learning techniques. In: 2018 International Conference on Artificial

- Intelligence and Data Processing (IDAP); 28-30 September 2018; Malatya, Turkey. 1–5.
- [17] Kolkur S, Kalbande D, Shimpi P, Bapat C, Jatakia J. Human skin detection using RGB, HSV and YCbCr color models. In: Proceedings of the International Conference on Communication and Signal Processing 2016 (ICCASP 2016); 20-25 March 2016; Sanghay, China. 324-332.
- [18] Gonzalez RC, Woods RE. Digital image processing. Prentice Hall, 2008.
- [19] Sergyán S. Color Content-based Image Classification. In: 5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics; 25 - 26 January 2007; Poprad, Slovakia. 427–434.
- [20] Bora D, Gupta AK, Khan F. Comparing the Performance of L\*A\*B\* and HSV Color Spaces with Respect to Color Image Segmentation. 2015. arXiv preprint arXiv:1506.01472.
- [21] Kotsiantis SB. Supervised machine learning: A review of classification techniques. *Informatica* 2007; 31: 249–268.
- [22] Shi L, Duan Q, Ma X, Weng M. The research of support vector machine in agricultural data classification. Springer, Berlin, Heidelberg 2012; 265–269.
- [23] Vapnik VN. The nature of statistical learning theory. Springer, 2000.
- [24] Cömert Z, Kocamaz AF, Subha V. Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment. *Comput. Biol. Med* 2018; 99: 85–97.
- [25] Cömert Z, Kocamaz AF. Comparison of machine learning techniques for fetal heart rate classification. *Acta Phys. Pol. A* 2017; 132(3): 451–454.
- [26] Cömert Z, Kocamaz AF. Open-access software for analysis of fetal heart rate signals. *Biomed. Signal Process. Control* 2018; 45: 98–108.