

**T.C.
İNÖNÜ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**HAPLOİD VE DİPLOİD MISIR TOHUMLARININ İMGE
ANALİZİ VE MAKİNE ÖĞRENME ALGORİTMALARI
KULLANILARAK SINIFLANDIRILMASI**

Yahya ALTUNTAŞ

**YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

Ağustos 2018

Tezin Başlığı : Haploid ve Diploid Mısır Tohumlarının İmge Analizi ve Makine Öğrenme Algoritmaları Kullanılarak Sınıflandırılması

Tezi Hazırlayan : Yahya Altuntaş

Sınav Tarihi : 09.07.2018

Yukarıda adı geçen tez jürimizce değerlendirilerek Bilgisayar Mühendisliği Ana Bilim Dalında Yüksek Lisans Tezi olarak kabul edilmiştir.

Sınav Jüri Üyeleri

Tez Danışmanı : **Dr.Öğr.Üyesi A. Fatih KOCAMAZ**
İnönü Üniversitesi

Doç.Dr. Fatih TALU
İnönü Üniversitesi

Prof.Dr. Abdulkadir ŞENGÜR
Fırat Üniversitesi

Prof.Dr. Halil İbrahim ADIGÜZEL
Enstitü Müdürü

ONUR SÖZÜ

Yüksek Lisans Tezi olarak sunduğum “*Haploid ve Diploid Mısır Tohumlarının İmge Analizi ve Makine Öğrenme Algoritmaları Kullanılarak Sınıflandırılması*” başlıklı bu çalışmanın bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın tarafımdan yazıldığını ve yararlandığım bütün kaynakların, hem metin içinde hem de kaynakça yöntemine uygun biçimde gösterilenlerden oluştuğunu belirtir, bunu onurumla doğrularım.

Yahya ALTUNTAŞ

ÖZET

Yüksek Lisans Tezi

HAPLOİD VE DİPLOİD MISIR TOHUMLARININ İMGE ANALİZİ VE MAKİNE ÖĞRENME ALGORİTMALARI KULLANILARAK SINIFLANDIRILMASI

Yahya ALTUNTAŞ

İnönü Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

59+ix sayfa

2018

Danışman: Dr.Öğr.Üyesi A. Fatih KOCAMAZ

Katlanmış haploid bitkiler günümüzde gelişmiş mısır ıslah programlarında yaygın olarak kullanılmaktadır. Bu teknik ıslah süresini kısaltmakta ve ıslahın etkinliğini arttırmaktadır. Bu ıslah tekniğinde önemli işlerden biri haploid tohumların seçimidir. Haploid tohumların seçiminde en yaygın yöntem R1-nj renk markörüdür. Bu seçimin elle yapılıyor olması seçim başarımını azaltmakta, zaman ve emek kaybına sebep olmaktadır. Bu nedenle seçim başarımını arttıracak, zaman ve emek tasarrufu sağlayacak otomatik seçim yöntemlerinin geliştirilmesi bir ihtiyaç haline gelmiştir.

Bu tez çalışmasında, R1-nj renk markörüne göre haploid ve diploid mısır tohumlarını imge analizi ve makine öğrenme algoritmaları kullanarak sınıflandıracak bir yöntem önerilmiştir. Tez çalışması kapsamında iki ayrı veri seti oluşturulmuştur. Öznitelik vektörü olarak 8, 16 ve 32 renk ölçekli gri seviyeli renk histogramları, renk momentleri ve doku öznitelikleri kullanılmıştır. Öznitelikler karar ağacı, k-en yakın komşu, destek vektör makinesi ve yapay sinir ağı yöntemleri ile sınıflandırılmıştır. Sınıflandırıcı başarımı 10-katlı çapraz doğrulama yöntemi ile sınanmıştır. En iyi performans, k-en yakın komşu sınıflandırma yöntemi ile renk momentleri öznitelikleri kullanılarak %90,39 doğruluk oranıyla elde edilmiştir.

ANAHTAR KELİMELEER: haploid tanıma, sınıflandırma, imge analizi, renk histogramları, renk momentleri

ABSTRACT

Master Thesis

CLASSIFICATION OF HAPLOID AND DIPLOID MAIZE SEEDS BY USING IMAGE ANALYSIS AND MACHINE LEARNING ALGORITHMS

Yahya ALTUNTAŞ

İnönü University

Graduate School of Natural and Applied Sciences

Department of Computer Engineering

59+ix pages

2018

Supervisor: Asst.Prof. A. Fatih KOCAMAZ

Doubled haploids is now widely used in advanced maize breeding programs. This technique shortens the breeding period and increases the efficiency of breeding. One of the important processes in this breeding technique is the selection of haploid seeds. The most common method of selecting haploid seeds is the R1-nj color marker. The fact that this selection is performed manually reduces the selection success and causes time and labor loss. For this reason, it has become a need to develop automatic selection methods that will save time and labor and increase selection success.

In this thesis, a method is proposed to classify haploid and diploid maize seeds using image analysis and machine learning algorithms according to the R1-nj color marker. Within the scope of the thesis study, two separate datasets were created. As the feature vectors; 8, 16 and 32 bin grayscale color histograms, color moments and texture features were used. The obtained feature vectors is classified by decision tree, k-nearest neighbors, support vector machine and artificial neural network. The classifier performance was tested by a 10-fold cross-validation method. Best performance has been obtained with 90.39% accuracy rate by using the color moments features with the k-nearest neighbor classification method.

KEYWORDS: haploid identification, classification, image analysis, color histograms, color moments

TEŞEKKÜR

Tez çalışmamın her aşamasında gerek bilgi ve tecrübesiyle, gerekse samimi ve insani yönüyle desteklerini esirgemeyen tez danışmanım Dr. Öğr. Üyesi Adnan Fatih KOCAMAZ'a;

Tezimin düzenlenmesi aşamasında yardımlarını esirgemeyen değerli arkadaşlarım Bitlis Eren Üniversitesi Dr. Öğr. Üyesi Zafer CÖMERT, İnönü Üniversitesi Öğr. Gör. Emrah DÖNMEZ, Bilgisayar Mühendisi Serdar ALASU ve Muş Alparslan Üniversitesi Arş. Gör. M. Murat TURHAN'a;

Veri setlerimizi temin ettiğimiz Sakarya Mısır Araştırma Enstitüsü Müdürlüğü ve veri setlerinin hazırlanmasında değerli destekleri için Dr. Rahime Cengiz ve Dr. Mesut Esmeray'a;

Ayrıca bu süreçte sabır ve anlayışından ötürü, her zaman yanımda olan sevgili eşim Meryem ALTUNTAŞ'a ve tüm aileme ve dostlarıma teşekkür ederim.

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
İÇİNDEKİLER	iv
SİMGELER VE KISALTMALAR	vi
ŞEKİLLER DİZİNİ	vii
ÇİZELGELER DİZİNİ	ix
1. GİRİŞ	1
1.1. Haploid Bitki Elde Etme Yöntemleri	1
1.2. In Vivo Maternal Haploid Tekniği	2
1.3. Haploid Seçimi	2
2. İLİŞKİLİ ÇALIŞMALAR	6
3. MATERYAL ve YÖNTEM	11
3.1. Veri Setleri	12
3.1.1. Mısır tohumları	12
3.1.2. Görüntülerin elde edilmesi	15
3.1.3. Mısır tohumlarının imgeden bölütlenmesi	17
3.2. Embriyo Bölütleme	19
3.3. Öznitelik Çıkarımı	21
3.3.1. Renk histogramları	22
3.3.2. Renk momentleri	28
3.3.3. Gri seviyeli eş oluşum matrisi	30
3.4. Sınıflandırma	33
3.4.1. Karar Ağaçları	34
3.4.2. k-en yakın komşuluk	35

3.4.3.	Destek vektör makinaları	36
3.4.4.	Yapay sinir ağıları	37
3.5.	Performans Ölçümleri	39
3.5.1.	k-Katlı Çapraz Doğrulama	39
3.5.2.	Hata Matrisi ve Performans Metrikleri	40
3.5.3.	AİK Eğrileri	42
4.	ARAŞTIRMA BULGULARI	44
4.1.	Sınıflandırma Sonuçları	44
4.1.1.	Renk histogramları	44
4.1.2.	Renk momentleri	49
4.1.3.	Doku öznitelikleri	51
5.	TARTIŞMA VE SONUÇ	53
	KAYNAKLAR	56
	ÖZGEÇMİŞ	59

SİMGELER VE KISALTMALAR

AİK	Alıcı işletim karakteristiği
AİK-AKA	Alıcı işletim karakteristiği altında kalan alan
DVM	Destek vektör makineleri
FAO	Gıda ve Tarım Örgütü
GSEM	Gri seviyeli eş oluşturma matrisi
KA	Karar ağaçları
k-EK	k-En yakın komşuluk
KH	Katlanmış haploid
KH2016	Katlanmış haploid 2016 veri seti
KH2017	Katlanmış haploid 2017 veri seti
TÜİK	Türkiye İstatistik Kurumu
YSA	Yapay sinir ağları

ŞEKİLLER DİZİNİ

Şekil 1.1	In vivo maternal haploid melezlemesinden elde edilebilecek tohumlar.....	3
Şekil 1.2	(a) Haploid ve (b) diploid mısır tohumlarında R1-nj renklenmesi	4
Şekil 3.1	Yapılan çalışmaların genel akış şeması.....	11
Şekil 3.2	Görüntüleme düzeneği.....	15
Şekil 3.3	KH2016 veri seti ham görüntü örneği.....	16
Şekil 3.4	KH2017 veri seti ham görüntü örneği.....	16
Şekil 3.5	Mısır tohumlarının imgeden bölütlenmesi aşamaları.....	18
Şekil 3.6	KH2016 veri setine ait (a) haploid ve (b) diploid mısır tohum örnekleri.....	19
Şekil 3.7	KH2017 veri setine ait (a) haploid ve (b) diploid mısır tohum örnekleri.....	19
Şekil 3.8	Endospermde R1-nj renklenmesi.....	20
Şekil 3.9	Tohum kütle merkezi ve öznitelik çıkarımı yapılan bölge.....	21
Şekil 3.10	Diploid tohum örneğinde öznitelik çıkartılan bölge ve renk histogramı.....	23
Şekil 3.11	Haploid tohum örneğinde öznitelik çıkartılan bölge ve renk histogramı.....	23
Şekil 3.12	Referans ve komşu pikseller arasındaki açısal ilişki.....	31
Şekil 3.13	GSEM oluşturulması.....	31
Şekil 3.14	Karar ağacı modeli.....	34
Şekil 3.15	DVM modeli.....	36
Şekil 3.16	Çok katmanlı örnek bir yapay sinir ağı.....	38
Şekil 3.17	k-katlı çapraz doğrulama.....	40
Şekil 3.18	Hata matrisi.....	41
Şekil 4.1	KH2016 veri seti 8 renk ölçekli gri seviyeli renk histogramı öznitelik vektörünün sınıflandırıcı AİK eğrileri.....	45
Şekil 4.2	KH2017 veri seti 8 renk ölçekli gri seviyeli renk histogramı öznitelik vektörünün sınıflandırıcı AİK eğrileri.....	46
Şekil 4.3	KH2016 veri seti 16 renk ölçekli gri seviyeli renk histogramı öznitelik vektörünün sınıflandırıcı AİK eğrileri.....	47
Şekil 4.4	KH2017 veri seti 16 renk ölçekli gri seviyeli renk histogramı öznitelik vektörünün sınıflandırıcı AİK eğrileri.....	47
Şekil 4.5	KH2016 veri seti 32 renk ölçekli gri seviyeli renk histogramı öznitelik vektörünün sınıflandırıcı AİK eğrileri.....	48
Şekil 4.6	KH2017 veri seti 32 renk ölçekli gri seviyeli renk histogramı öznitelik vektörünün sınıflandırıcı AİK eğrileri.....	49
Şekil 4.7	KH2016 veri seti renk momentleri öznitelik vektörünün sınıflandırıcı AİK eğrileri.....	50

Şekil 4.8	KH2017 veri seti renk momentleri öznitelik vektörünün sınıflandırıcı AİK eğrileri.....	50
Şekil 4.9	KH2016 veri seti doku öznitelik vektörünün sınıflandırıcı AİK eğrileri.....	52
Şekil 4.10	KH2017 veri seti doku öznitelik vektörünün sınıflandırıcı AİK eğrileri.....	52
Şekil 5.1	Yanlış sınıflandırılmış (a) haploid ve (b) diploid mısır tohum örnekleri.....	54



ÇİZELGELER DİZİNİ

Çizelge 3.1	KH2016 verisetinde kullanılan donör genotipler.....	13
Çizelge 3.2	KH2017 verisetinde kullanılan donör genotipler.....	14
Çizelge 3.3	KH2016 veri seti 8 renk ölçekli gri seviye renk histogram özniteliklerinin aldığı değerler.....	24
Çizelge 3.4	KH2017 veri seti 8 renk ölçekli gri seviye renk histogram özniteliklerinin aldığı değerler.....	24
Çizelge 3.5	KH2016 veri seti 16 renk ölçekli gri seviye renk histogram özniteliklerinin aldığı değerler.....	25
Çizelge 3.6	KH2017 veri seti 16 renk ölçekli gri seviye renk histogram özniteliklerinin aldığı değerler.....	25
Çizelge 3.7	KH2016 veri seti 32 renk ölçekli gri seviye renk histogram özniteliklerinin aldığı değerler.....	26
Çizelge 3.8	KH2017 veri seti 32 renk ölçekli gri seviye renk histogram özniteliklerinin aldığı değerler.....	27
Çizelge 3.9	KH2016 veri seti renk momentleri özniteliklerinin aldığı değerler.....	29
Çizelge 3.10	KH2017 veri seti renk momentleri özniteliklerinin aldığı değerler.....	30
Çizelge 3.11	KH2016 veri seti doku özniteliklerinin aldığı değerler.....	33
Çizelge 3.12	KH2017 veri seti doku özniteliklerinin aldığı değerler.....	33
Çizelge 4.1	Değerlendirme metrikleri.....	44
Çizelge 4.2	8 renk ölçekli gri seviyeli renk histogramı sınıflandırma sonuçları.....	45
Çizelge 4.3	16 renk ölçekli gri seviyeli renk histogramı sınıflandırma sonuçları.....	46
Çizelge 4.4	32 renk ölçekli gri seviyeli renk histogramı sınıflandırma sonuçları.....	48
Çizelge 4.5	Renk momentleri öznitelik vektörü sınıflandırma sonuçları....	49
Çizelge 4.6	Doku öznitelik vektörü sınıflandırma sonuçları.....	51

1. GİRİŞ

Mısır (*Zea mays* L.) gerek insan ve hayvan beslenmesinde, gerekse sanayide kullanılan önemli bir bitkidir. Orta Amerika ve Sahra Altı Afrika'sında bir milyardan fazla insanın temel gıdasıdır [1]. Gıda ve Tarım Örgütü (FAO) 2016 verilerine göre dünya mısır üretim miktarı 1.060.107.470 ton ile şeker kamışı üretiminden sonra ikinci sırada gelmektedir [2]. Türkiye İstatistik Kurumu (TÜİK) 2017 verilerine göre Türkiye silajlık ve dane mısır üretim miktarı 29.052.841 ton ile ilk sırada yer almaktadır [3].

Artan dünya nüfusu ve iklim değişiklikleri yüksek verimli, biyotik ve abiyotik strese dayanıklı yeni çeşitlerin geliştirilmesini zorunlu hale getirmiştir. Aranılan özelliklerde yeni çeşitler geliştirmek ancak mısır ıslah programlarıyla mümkün olabilmektedir. Mısır ıslah çalışmaları ile talep edilen özelliklerde yeni çeşitler geliştirmek amaçlanmaktadır.

Mısır ıslah programlarında temel aşama melez genotiplere ebeveyn olabilecek saf hatların geliştirilmesidir. Geleneksel ıslah çalışmalarıyla bu saf hatların elde edilmesi 6-7 yıl sürmektedir. Bu sürenin sonunda ancak %99 homozigotluk düzeyine ulaşılabilmektedir. Bu sürenin kısaltılmasında haploid bitki elde etme teknikleri önemli avantajlar sağlamaktadır [4]. % 100 homozigot ve homojen hatların hızlı gelişimi, katlanmış haploid (KH) hatlarının geliştirilmesi ile gerçekleştirilmektedir [5].

İlk haploid mısır bitkileri Stadler ve Randolph (1929 ve 1932) tarafından tanımlanmıştır. Yaklaşık yirmi yıl sonra, Chase (1947), Amerika mısır kuşağı materyallerinde düşük oranda (1/1000) haploidler bulmuştur. Daha sonraları kromozom ikiye katlama maddesi olarak kolhisin, ve *in vivo* haploid indirgeme için uygun spesifik genotiplerin saptanmasıyla haploid bitkilerin melez ıslah programlarında kullanımı yaygınlaşmıştır. Bu gelişmeler son yirmi yılda *in vivo maternal haploid* sistemi kullanarak hat geliştirmede önemli bir artışa yol açmıştır [5].

1.1. Haploid Bitki Elde Etme Yöntemleri

Haploid bitkiler "*in vitro*" ve "*in vivo*" olarak elde edilebilmektedir. *In vitro* haploid bitki elde etme teknikleri laboratuvar şartlarında uygulanmaktadır. Ancak, *in vitro* uygulamalarında genotip etkisi sebebiyle arzu edilen düzeyde haploid bitki elde edilememektedir [4].

Ticari olarak geliştirilen mevcut katlanmış hatların birçoğunun *in vivo* haploid tekniği ile elde edildiği, diğer tekniklerin ise katlanmış hat geliştirmede daha az etkili olduğu bildirilmektedir [5]. *In vivo* haploid bitki elde etme tekniğinde indirgeyici hat olarak adlandırılan hatlar kullanılmaktadır. *In vivo* tekniğinde indirgeyici hatlar tozlayıcı olarak kullanılmakta ve toz verdiği bitkinin koçanlarında haploid olan tohumların oluşmasına imkan vermektedir. Doğada çok düşük frekansta (1/1000) [5] bulunan haploidler, indirgeyici hatlar kullanılarak daha yüksek frekanslarda elde edilebilmektedir. İndirgeyici hattın tozlayıcı olarak kullanıldığı yöntem maternal haploidi, toz alıcı olarak kullanıldığı yöntem paternal haploidi adı verilmektedir. Maternal haploid yöntemiyle, paternal haploid yöntemine göre elde edilen haploid oranı daha yüksek olmaktadır [6]. *In vivo* maternal haploid tekniği ile haploid bitki elde etme başarısı kullanılan indirgeyici hattın özelliklerine göre değişmekle birlikte %2-15 arasındadır [7]. *In vivo* maternal haploid tekniği yüksek haploid bitki elde etme başarısı nedeniyle en yaygın kullanılan tekniktir.

1.2. *In Vivo* Maternal Haploid Tekniği

In vivo maternal haploid tekniğinde indirgeyici hat ile donör bitki melezlenir ve haploidler üretilir. Melezleme sonucunda elde edilen haploid ve diploid bireyler ayrıştırılır. Son olarak, haploid bitkiler, kromozom sayısını ikiye katlamak için kolhisin ile muamele edilir. Kolhisin ile muamele edilmiş haploidler kendilenerek nihai %100 homozigotlukta KH hatlar elde edilir.

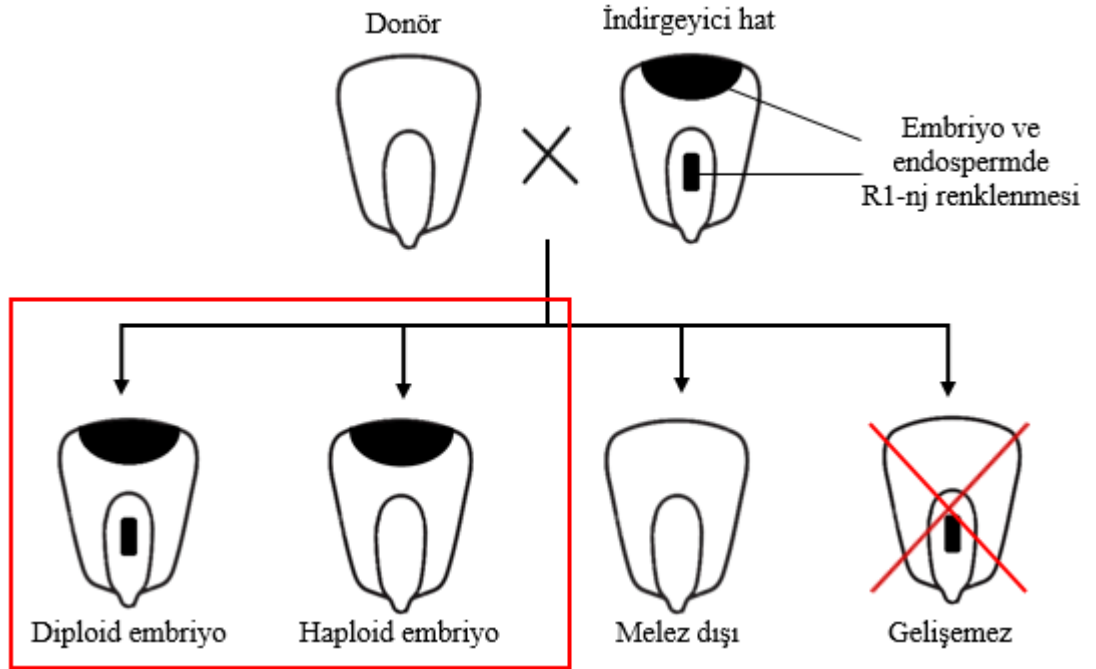
Kolhisin muamelesinin sadece donör bitkinin kromozomlarını içeren haploidlere uygulanması gerekir. Çünkü diploidler hem donör hem de tozlayıcı bitkilerin kromozomlarını içerir ve $2n$ kromozoma sahiptirler. Tohum döneminde haploidlerin diploidlerden ayrılması; suni kromozom katlamasına izin vermek, maliyeti düşürmek, işçiliği azaltmak, sera çalışmalarını ve tarla alanını azaltmak gibi avantajlar sağlar [8].

1.3. Haploid Seçimi

Haploid ve diploid mısır tohumlarının embriyo ağırlıkları arasındaki fark [9] ve farklı yağ içeriği düzeyleriyle [10] sınıflandırma çalışmaları yapılsa da, en yaygın ve başarılı biçimde kullanılan seçilebilir işaretleyici R1-Navajo (R1-nj) renk markörüdür. R1-nj renk markörü ilk olarak Nanda and Chase [11] tarafından

kullanılmıştır. İndirgeyici (tozlayıcı) hattın sahip olduğu baskın bir gen olan R1-nj embriyo ve endospermde mor-kırmızı renklenme olarak gözlemlenir. Haploid bitkiler tek n kromozomlu embriyoya sahiptirler ve sadece anne ebeveynin kromozomlarını taşıdıkları için embriyolarında R1-nj renklenmesi gözlenmez. Bu sayede, R1-nj renk markörü haploid ve diploid tohumların görsel sınıflandırmasına izin verir.

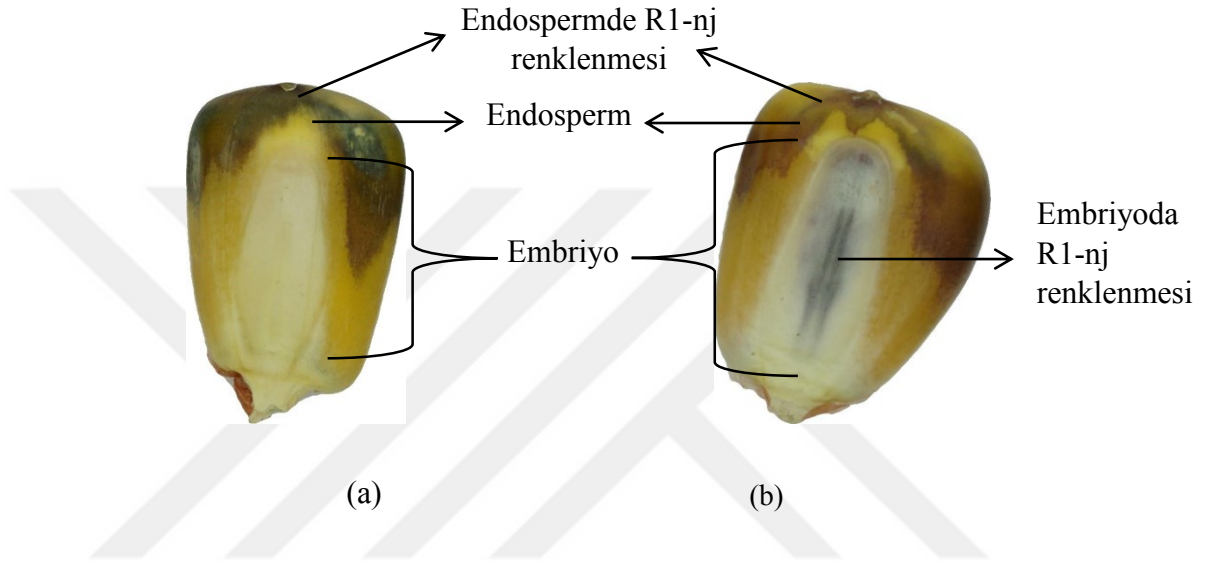
In vivo maternal haploid indüksiyondan elde edilen tohumların embriyo ve endospermdeki renk durumu üç şekilde ortaya çıkabilir. Embriyo ve endospermde renklenme olmayan tohumların kontamine oldukları varsayılır ve melez dışı olarak isimlendirilirler. Hem embriyo hem de endospermde kırmızı-mor renklenme gözlemlenen tohumlar hem donör hem de indirgeyici hattın kromozomlarını içerdiklerinden 2n kromozoma sahiptirler ve diploid olarak isimlendirilirler. Endospermde R1-nj renklenmesi oluşmasına rağmen, renksiz embriyoya sahip olan tohumlar genetik olarak sadece donör bitkinin kromozomlarını taşırlar ve tek n kromozoma sahiptirler ve haploid olarak isimlendirilirler. Şekil 1.1 [5]'den esinlenilerek oluşturulmuştur ve *in vivo* maternal haploid melezlemesi sonucu elde edilebilecek tohumları göstermektedir.



Şekil 1.1 *In vivo* maternal haploid melezlemesinden elde edilebilecek tohumlar

Şekil 1.2'de diploid ve haploid mısır tohum örnekleri gösterilmiştir. R1-nj'nin embriyo ve endospermdeki farklı ifadeleri haploid ve diploid mısır tohumlarının görsel

olarak sınıflandırılabilmesini sağlar. Ancak, haploidlerin R1-nj markör sistemine göre elle seçimi zaman alan, emek yoğun bir süreçtir. R1-nj renklenmesi küçük bir noktadan, tüm endospermi kapsayan renkliliğe kadar çeşitlilik gösterir. Ayrıca rengin koyuluğu da çok açıktan koyuya ve daha derine doğru değişiklik gösterir. R1-nj renklenmesindeki bu çeşitlilik elle yapılan sınıflandırmada yüksek hata oranına neden olur. Bu sebeple, haploid tohumların seçimini otomatik olarak gerçekleştirecek bir sisteme gereksinim duyulmaktadır.



Şekil 1.2 (a) Haploid ve (b) diploid mısır tohumlarında R1-nj renklenmesi

Bu tez çalışmasında, haploid ve diploid mısır tohumlarının sınıflandırılması amaçlanmıştır. Çalışma kapsamında biri 413 örnekten, diğeri 3.000 örnekten oluşan iki farklı veri seti hazırlanmıştır. Veri setlerini oluşturan örneklerde tohum embriyo bölgeleri öznitelik çıkarımı yapmak üzere bölütlenmiştir. Bölütlenen embriyo bölgelerinden renk ve doku öznitelikleri çıkarılmıştır. Renk öznitelikleri olarak renk histogramları ve renk momentleri; doku özniteliği olarak gri seviyeli eş oluşum matrisi kullanılmıştır. Renk histogram öznitelik vektörleri 8, 16 ve 32 renk ölçekli gri seviye embriyo imgelerinin renk histogram değerleri kullanılarak çıkarılmıştır. Renk momentleri öznitelik vektörü RGB renk uzayında embriyo imgelerinin her renk kanalından ayrı ayrı olmak üzere ilk dört derece renk momentleri kullanılarak oluşturulmuştur. Doku öznitelik vektörü embriyo imgelerinin gri seviye eş oluşum matrisinden elde edilen enerji, homojenlik, kontrast, korelasyon ve entropi değerlerinden oluşturulmuştur. Sınıflandırıcı olarak kullanılan karar ağaçları, k-en yakın komşuluk sınıflandırıcısı, destek vektör makineleri ve yapay sinir ağları her veri

seti için beşer tane olmak üzere toplam on öznitelik vektörü ile eğitilmiştir. 10-katlı çapraz doğrulama yöntemi ile sınıflandırıcı performansları test edilmiştir. Elde edilen deney sonuçlarına göre en iyi performans, k-en yakın komşu sınıflandırma yöntemi ile renk momentleri öznitelikleri kullanılarak %90,39 doğruluk oranıyla elde edilmiştir.

Haploid tohumların otomatik seçimini gerçekleştirmeyi amaçlayan bazı pilot çalışmalara bölüm 2’de yer verilmiştir. Bölüm 3’de bu çalışma kapsamında kullanılan veri seti ve imge analizi, öznitelik çıkarımı ve seçimi ile sınıflandırma yöntemleri ve sınıflandırıcı performanslarının karşılaştırılmasında kullanılan metrikler açıklanmıştır. Veri setleri üzerinde uygulanan yöntemlerden elde edilen deney sonuçları Bölüm 4’de verilmiştir. Bölüm 5’de deney sonuçları tartışılmış ve elde edilen sonuçlar değerlendirilmiştir.

2. İLİŞKİLİ ÇALIŞMALAR

Haploid ve diploid mısır tohumlarının seçimi için farklı teknikler olmakla birlikte, en yaygın ve başarılı biçimde kullanılan seçilebilir işaretleyici R1-nj renk markörüdür. R1-nj renk markörü haploid ve diploid mısır tohumlarının görsel ayırımına izin verir. Bu seçim işleminin elle yapılıyor olması, seçim başarımı azaltmakta, zaman ve emek kaybına sebep olmaktadır. Günümüzde ise imge işleme ve makine öğrenme yöntemleri kullanılarak haploid mısır tohumlarının sınıflandırma işlemi için otomatik sistemlerin geliştirilmesi amaçlanmaktadır. Otomatik sistemler hız, dayanıklılık ve doğruluk prensiplerine uygun olacak şekilde geliştirilmektedir. Bu çalışma kapsamında otomatik sistem geliştirmeye yönelik yapılan araştırma ve çalışmalardan en önemli olanlarına değinilmiştir.

Wang vd. [12] çalışmalarında, LSTM-CNN algoritmasının optimum dalga bandı seçimine dayalı, haploid ve diploid mısır tohumlarının yüksek doğruluklu bir tanımlama modelini derin öğrenme ve hiperspektral görüntüleme teknolojisi ile belirlenen optimum 1367.6-1526.4nm dalga bandında %97'ye ulaşan doğrulukla gerçekleştirilmiştir. Deneysel örnekler olarak 100 haploid ve 100 diploid olmak üzere 200 tohum kullanmışlardır. Yapılan çalışmada, önerilen modelin mısır haploid tanımda olağanüstü bir etkiye sahip olduğu ve bir dereceye kadar genelleştirilebileceği iddia edilmiştir.

Lin vd. [13] yaptıkları çalışmalarında, yakın kızıl ötesi (NIR) spektroskopi yaygın iletim (DT) teknolojisi sayesinde, mısır haploid tohumlar NIR spektrum özellikleri kullanılarak otomatik olarak seçilebileceğini söylemişlerdir. Ancak bu tür bir sistemde mısır tohumlarının NIR spektrumları, tanımlama performansını bozacak çok sayıda fazlalık ve gürültü içereceğini de belirtmişlerdir. Bu sorunun üstesinden gelmek için, toplanan spektrumları iyileştirecek bir düşük boyutlu ve tohum spektrum özelliklerinin tekdüze bir uzayını tasarlamışlardır. Sıfır-fazlı bileşen analizi (ZCA) yöntemi, özellik uzayını eşitlemek için kullanmışlar ve düşük boyut uzayını tasarlamak için kısmi en küçük kareler regresyonunu (PLSR) kullanmışlardır. Daha sonra, geri yayılma sinir ağının (BPNN) sınıflandırıcı kullanarak, mısır haploit tohumlarının seçilmesi için yüksek niteliksel bir tanımlama yöntemi geliştirilmişlerdir. Gerçekleştirdikleri deneyler neticesinde önerilen yöntemin ortalama doğruluğunun %96.16 civarında olduğunu öne sürmüşlerdir.

Song vd. [14] haploid mısır tohumlarını diploid mısır tohumlarından ayırabilen bir otomatik sınıflandırma sistemi geliştirmişlerdir. Sistemlerinde üç temel birimin olduğunu belirtmişlerdir. Tohum besleme birimi, mısır tohumlarını senkronize bant üzerinden dağıtır. Görüntü elde etme birimi, mısır tohumunun görüntülerini elde etmenin yanı sıra diploidlerin, haploid tohumlardan, embriyonun renk özelliğine göre ayrılmasını da sağlar. Ayırma ünitesi ise hava emişini kullanarak diploid tohumları seçebilen mekanik kollar ve solenoid valfler içerir. Gerçekleştirdikleri deneyler sonucunda veri seti içerisinde diploid tohumlar için tanımlama doğruluğu oranlarının %90 üzerinde başarıldığını iddia etmektedirler.

Fuente vd. [15] VideometerLab 3 yazılımının performansını değerlendirmek, elle sınıflandırmada zorluk yaşanan farklı genotipleri sistemde test etmek ve haploid-diploid ayrımını doğrulamak için DNA işaretleme analizlerini kullanmayı amaçlamışlardır. Haploid mısır tohumlarını ayırt etmek için önerdikleri yöntemde altı kendilenmiş hat ile maternal haploid indükleyici 'RWS / RWK-76' ile çaprazlamış ve her bir hat için bir tohum örneğini elle ayırmışlardır. Kullandıkları veri-setinde altı kendilenmiş hat için analiz edilen 100 haploid ve 100 diploid olmak üzere toplamda 200 tohum bulunmaktadır. VideometerLab 3 sistemini kullanarak, haploidler ve diploidler arasında ayırım yapmak için spektral görüntüleme teknikleri uyguladılar. Çalışmalarında test edilmiş tohumun haploid/diploid durumunu doğrulamak için DNA işaretleyicileri kullanarak, genotiplerin çoğunluğu için haploid tanımlamasının % 50'den fazla doğrulukla mümkün olduğunu söylemişlerdir.

Boote vd. [16] mısır tohumlarında haploid seçimi için yeni bir flüoresan bazlı yöntem, floresan mikrospektroskopi ve görüntüleme ile birlikte R1-nj renk işaretleyicisi kullanılarak geliştirmiştir. Çalışmada değişen R1-nj ekspresyonuna sahip yedi kendilenmiş hat kullanıldı. Embriyonik boya bölgesinde diploid tanelerin floresans tepkisinin aynı anda düşük yoğunluk sergilediği ve boyasız bölge haploid embriyoların floresanından daha yüksek bir dalga boyunda ortaya çıktığını göstermişlerdir. Yoğun örnek popülasyonlarından haploidleri ayırmak için floresan görüntülere yoğunluk ve alan eşikleri uyguladılar ve yedi doğal yolun tamamında %80'den fazla sıralama verimliliği elde etmişlerdir. Floresan görüntülemenin tohum işleme için mevcut teknolojiler ile kombine edildiğinde yüksek sınıflandırma verimliliği ile gelecek vaat eden bir yöntem olduğunu ifade etmişlerdir.

Qin vd. [17] haploid mısır tohumu tanımlama kapsamında yaptıkları araştırmalarda MicroNIR-1700 serisi minyatür yakın kızılötesi spektrometre cihazını kullanmışlardır. Yakın Kızılötesi Spektroskopisi (NIRS) nitel analiz tekniklerine dayanarak, haploid mısır tohumlarını tanımlamak için yansıma ve geçirgenlik spektrumları kullanılarak karşılaştırmalı bir çalışma gerçekleştirdiklerini söylemişlerdir. Kısmi en küçük kareler-diskriminant analizi (PLS-OLDA) ön-işlemden geçirilmiş spektral verileri sıkıştırmak için kullanılmış ve daha sonra tanımlama modelleri destek vektör makinesi temel alınarak oluşturulmuştur. Ölçülen veriler yansıma ve geçirgenlik modlarında kaydedilmiş ve tanıma doğruluğu oranları hesaplanmıştır. Yürütülen deneyler neticesinde yansıma modunda alınan ölçümler için, ortalama tanıma oranı, embriyo tarafı pozisyonuna bakılmaksızın % 60 altında kaldığını, geçirgenlik modunda ise, ortalama tanıma oranı % 93,2'ye ulaştığını belirtmişlerdir. Deneylerden yola çıkarak, geçirgenlik spektrumları ile haploid mısırlarının etkili bir şekilde tanımlanabileceğini ve tohum pozisyonlarına duyarsızlaşabileceğini iddia etmişlerdir.

Zhang vd. [18] haploid tohumların manuel olarak seçiminin zor olduğunu, seçim verimliliğini düşürdüğünü ve iş, zaman tasarrufu sağlayacak yüksek doğruluğa sahip bir yöntemin gerekliliğini belirtmişlerdir. Mısır embriyosunun özelliklerinin haploid ve hibrid tohumlar arasında ayırım yapmak için önemli olduğunu vurgulamışlardır. Bu kapsamda haploid ve diploid mısır tohumlarını sınıflandırmak için embriyo üzerindeki R1-nj renk markörüne makine görmesi yöntemlerini kullanarak özellik çıkarımını yapmak yoluyla dinamik tanıma yöntemi geliştirmişlerdir. RGB renk uzayındaki B kanalının ortalama piksel değerlerini kullanarak embriyo karakteristiklerini çıkarmak için bir görüntü segmentasyon yöntemi önermişlerdir. Navajo genetik işaretçileri ile mısır tohumlarının embriyo görüntüleri çıkarmışlardır. Elde edilen embriyo bölgesindeki Navajo genetik belirteçlerini RGB renk uzayında çıkarmış ve her bir tohum için Navajo piksellerinin sayısını tespit etmek yoluyla hangi sınıflandırmaya ait olduğunu belirlenmiştir. Navajo işaretçileri ile çıkarımı yapılan mısır embriyo özelliklerini kullanan hızlı bir haploid tohumlarını tanıma algoritması geliştirdiklerini iddia etmişlerdir. Gerçekleştirdikleri deneylerde yöntemin haploid tohumları %98.04, diploid tohumları ise %94.44 oranında tanıdığını belirtmişlerdir. Ancak çok az sayıda ve ön seçim yapılmış tohum

kullanarak deneyleri gerçekleştirdikleri için daha fazla tohum örneklemesinde iddia edilen başarıma ulaşma durumu belirsizdir.

Couto vd. [19] akış sitometrisi, morfolojik özellikler ve moleküler işaretçileri kullanarak haploid mısırların tanınmasına yönelik bir çalışma gerçekleştirmişlerdir. Haploidleri akış sitometrisi ile tanımlamak ve nükleer DNA içeriğini, bunlara neden olan tohumların morfolojik ve morfometrik özellikleri ile ilişkilendirmişlerdir. Haploidin androjenik doğasını doğrulamak için moleküler işaretçiler kullandılar. Elde edilen tohumları, kendilenmiş hat W23 ve ticari hibrid P30F90 arasındaki çaprazlamadan elde ettiler. Bu tohumlar arasından, bir grup olan embriyo beyaz ve perikarp morumsu olduğu varsayılan haploidler seçilmiştir. 330 tohumdan oluşan bu grup, tohum morfolojisi, tohum morfometrisi ve nükleer DNA içeriği temel alınarak karakterize edilmiştir. Elde ettikleri sonuçlara göre, kendilenmiş W23 hattının tropik koşullar altında bile haploid mısırları uyardığını iddia etmişlerdir. Mikrosatellit moleküler belirleyicilerin (SSR), haploidlerin androjenik özelliğini doğrulamada etkin olduğunu öne sürmüşlerdir.

Song vd. [20] çalışmalarında haploid ve diploid mısır tohumlarının özellik çıkarımı ve tanımlama yöntemi geliştirmişlerdir. Mısır taneleri, mısır haploid ıslah işleminde erkek ebeveynlerine gen işaretleyicileri ithal edildikten sonra hibrid indüksiyon ile üretilmiştir. Önerilen yöntem, mısır haploid çekirdeklerinin doğru tanımlanmasını gerçekleştirmiştir. Melez indüksiyon ile üretilen çekirdekleri araştırmışlardır. Yöntemlerinde ilk olarak, mısır tanelerini mor işaretler, sarı endosperm ve beyaz embriyo da dahil olmak üzere renk dağılımına göre üç parçaya ayırdılar. Bu üç parçadan alınan piksel örneklerinin özelliklerini renk analiz yöntemi ile çıkarmak için $(G - B) * R/B$ modelini önerdiler. Bu modelin, mor alanların etkili bir şekilde çıkarılmasını sağlayabileceğini vurgulamışlardır. Mısır tanelerinin morfolojik özellikleri ve mor işaretlerin dağılım özellikleri üzerine araştırma yaptıktan sonra, mısır haploid çekirdeklerinin basit ve hızlı bir tanımlama yöntemi oluşturduklarını iddia etmişlerdir. Deneyler sonucunda tanıma başarısının %98.07 olduğunu ve sitemlerinin dinamik test gerekliliklerini sağladığını öne sürmüşlerdir. Bu çalışmada kullanılan veri seti hakkında herhangi bir bilgi olmaması çalışmanın başarımının doğrulanması için yetersiz kalmasına neden olmaktadır.

Zhang vd. [21] BP sinir ağı kullanarak haploid mısır tohumlarının 1050-37 mısırın genetik işaretlerle araştırılması için bir görüntü bölütleme yöntemi

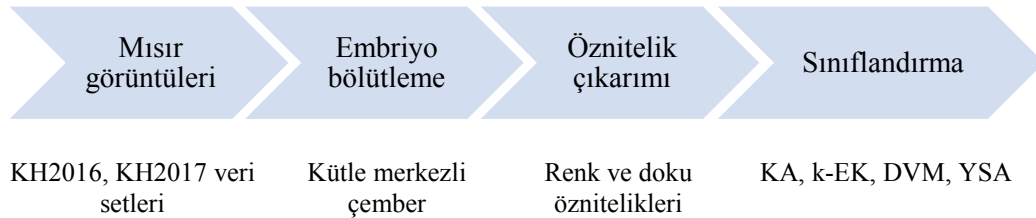
önermişlerdir. Renk özelliklerine göre, mısır tohumu görüntüleri mor alan, sarı alan ve beyaz alan olmak üzere üç renk desenine ayrılmıştır. Normalize rgb ve HSV renk uzayının farklı renk özellikleri analiz edilmiş ve mısır haploid tohumlarının etkili görüntü segmentasyonunu elde etmek için 3 katman içeren bir BP sinirsel ağ modeli oluşturmak için 7 özellik girdi parametreleri olarak seçilmiştir. Deneyler, modelin sınıflandırma doğruluklarının sırasıyla mor leke alanı için %97.61, sarı alan için %93.34 ve beyaz alan için %94.09 olduğunu göstermektedir. Deney sonuçlarına dayanarak BP NN tarafından elde edilen mor leke alanının, haploid çekirdeklerin ve hibrit çekirdeklerin tanımlanması için etkili ve güvenilir olduğunu söylemişlerdir.

Song vd. [22] bilgisayar görmesi yöntemlerini kullanarak haploid mısır tohumlarını seçme sistemi geliştirmişlerdir. Bu kapsamda Navajo işaretçileri ile işaretlenmiş mısır haploid çekirdekleri için bir sınıflandırma sistemi tasarlamışlardır. Geliştirdikleri sistem bir tohum iletim ünitesi, bir görüntü elde etme ve işleme birimi, bir sınıflandırma-boşaltma birimi ve bir sistem kontrol biriminden oluşmaktadır. Görüntü alma ve işleme birimi, haploid çekirdekleri, mısır embriyosu ve mısır endosperminin üstündeki renk özelliklerine göre mısır haploidlerini tanımlamıştır. Geliştirdikleri sistemin diploid tohumları haploid tohumlardan hava emiş kullanarak ayırdığını ve dakikada 500 tohumluk bir sınıflandırma hızı elde ettiklerini iddia etmişlerdir.

Song vd. [23] mısır tohumlarının dinamik kalite kontrolü ve derecelendirmesini sağlayan sistemi geliştirmiş, mısır tanelerinin dört sınıfa ayrılması için morfolojik özellikler kullanmış ve mısır tanelerini üç seviyeye ayırt etmek için renk özellikleri kullanılmışlardır. Ortalama sistem derecelendirme oranları sırasıyla %81.8 ve %93.04 olarak elde edilmiştir. Örüntü tanımaya dayalı mısır çeşitleri tanıma sistemi geliştirilmişlerdir. Sistem Nongda 80, Nongda 108, Gaoyou115, Nongda 4967 ve Bainuo 6'yı %92'lik ortalama tanıma doğruluğu ile mısır tohumlarını ayırt etmiştir. Bir haploid mısır tane derecelendirme sistemi geliştirilmiş, haploid mısır tohumları renk özellikleri ve örüntü tanıma teknolojisi ile tanınmıştır. İki dereceli serbestliğe sahip bir robot sistemi ile de sınıflandırma doğruluğunu %80 oranında sağladıklarını iddia etmektedirler.

3. MATERİYAL ve YÖNTEM

Bu tez çalışmasında, *in vivo* maternal haploid tekniği ile elde edilen haploid ve diploid mısır tohumlarının R1-nj markör sistemine göre imge analizi ve makine öğrenme algoritmaları kullanılarak sınıflandırılması amaçlanmıştır. Çalışma kapsamında biri 2016 yılında hasat edilen 413 haploid ve diploid mısır tohumundan oluşan, diğeri 2017 yılında hasat edilen 3.000 haploid ve diploid mısır tohumundan oluşan iki farklı veri seti hazırlanmıştır. Veri setlerini oluşturan örneklerde tohum embriyo bölgeleri öznitelik çıkarımı yapmak üzere bölütlenmiştir. Bölütlenen embriyo bölgelerinden renk ve doku öznitelikleri çıkarılmıştır. Renk öznitelikleri olarak renk histogramları ve renk momentleri, doku özniteliği olarak gri seviyeli eş oluşum matrisi kullanılmıştır. Renk histogram öznitelik vektörleri 8, 16 ve 32 renk ölçekli gri seviye embriyo imgelerinin renk histogram değerleri kullanılarak ayrı ayrı öznitelik vektörü olarak çıkarılmıştır. Renk momentleri öznitelik vektörü RGB renk uzayında embriyo imgelerinin her renk kanalından ayrı ayrı olmak üzere ilk dört derece renk momentleri kullanılarak oluşturulmuştur. Doku öznitelik vektörü embriyo imgelerinin gri seviye eş oluşum matrisinden elde edilen enerji, homojenlik, kontrast, korelasyon ve entropi değerlerinden oluşturulmuştur. Sınıflandırıcı olarak kullanılan karar ağaçları, k-en yakın komşuluk sınıflandırıcısı, destek vektör makineleri ve yapay sinir ağları her veri seti için beşer tane olmak üzere toplam on öznitelik vektörü ile eğitilmiştir. 10-katlı çapraz doğruluk yöntemi ile sınıflandırıcı performansları test edilmiştir. Elde edilen deney sonuçlarına göre, en iyi performans k-en yakın komşu sınıflandırma yöntemi ile renk momentleri öznitelikleri kullanılarak %90,39 doğruluk oranıyla elde edilmiştir. Tez çalışması kapsamında uygulanan tüm imge analizi ve sınıflandırma yöntemleri MATLAB R2016a ortamında gerçekleştirilmiştir. Yapılan çalışmaların genel akış şeması Şekil 3.1’de gösterilmiştir



Şekil 3.1 Yapılan çalışmaların genel akış şeması

3.1. Veri Setleri

İlgili yayınlarda araştırmacıların [12-23] haploid ve diploid mısır tohumu ayrımı için kullandıkları veri setleri sadece yayına özel olarak ilgili araştırmacılar tarafından oluşturulmuştur. Hali hazırda bu veri setleri genel kullanıma açık olmadıkları için bu tez çalışmasında 2 ayrı veri seti oluşturulmuştur. 2016 yılında hasat edilen mısır tohumlarından oluşan veri seti KH2016, 2017 yılında hasat edilen mısır tohumlarından oluşan veri seti KH2017 olarak isimlendirilmiştir. Tez çalışması kapsamında oluşturulan veri setleri detayları ile aşağıda anlatılmaktadır.

3.1.1. Mısır tohumları

Her iki veri setini oluşturan haploid ve diploid mısır tohumları Mısır Araştırma Enstitüsü Müdürlüğüne yürütülmekte olan “Ülkesel Mısır Islah Araştırmaları” projesi kapsamında *in vivo maternal haploid* indirgeme ile elde edilmiştir. Her iki veri setini oluşturan tohumların elde edilmesinde RWS, RWK-76 ve RWS x RWK-76 indirgeyici hatları kullanılmıştır. Donör olarak kullanılan F1 ve F2’ler FAO 450-700 arasında olum gruplarına, sarı atdışi ve waxy tane tiplerine sahiptir. Haploid ve diploid tohumların seçiminde R1-nj renk markörü dikkate alınmıştır. Verisetini oluşturan haploid ve diploid mısır tohumlarının seçiminde R1-nj renk geninin farklı ifadeleri olmasına (açık-koyu renk, yoğun-az renk gibi) dikkat edilmiştir.

KH2016 veri seti 87’si haploid, 326’sı diploid olmak üzere toplam 413 adet atdışi (*Zea mays indentata*) mısır tohumundan oluşmaktadır. Bu tohumlar, Çizelge 3.1’de listelenen 150 farklı donör genotip ile indirgeyici hatlar arasında yapılan indükleme melezlemesi sonucu elde edilmiştir.

KH2017 veri seti 1.230’si haploid, 1.770’si diploid olmak üzere toplam 3.000 adet atdışi (*Zea mays indentata*) mısır tohumundan oluşmaktadır. Bu tohumlar, Çizelge 3.2’de listelenen 107 farklı donör genotip ile indirgeyici hatlar arasında yapılan indükleme melezlemesi sonucu elde edilmiştir.

Çizelge 3.1 KH2016 verisetinde kullanılan donör genotipler

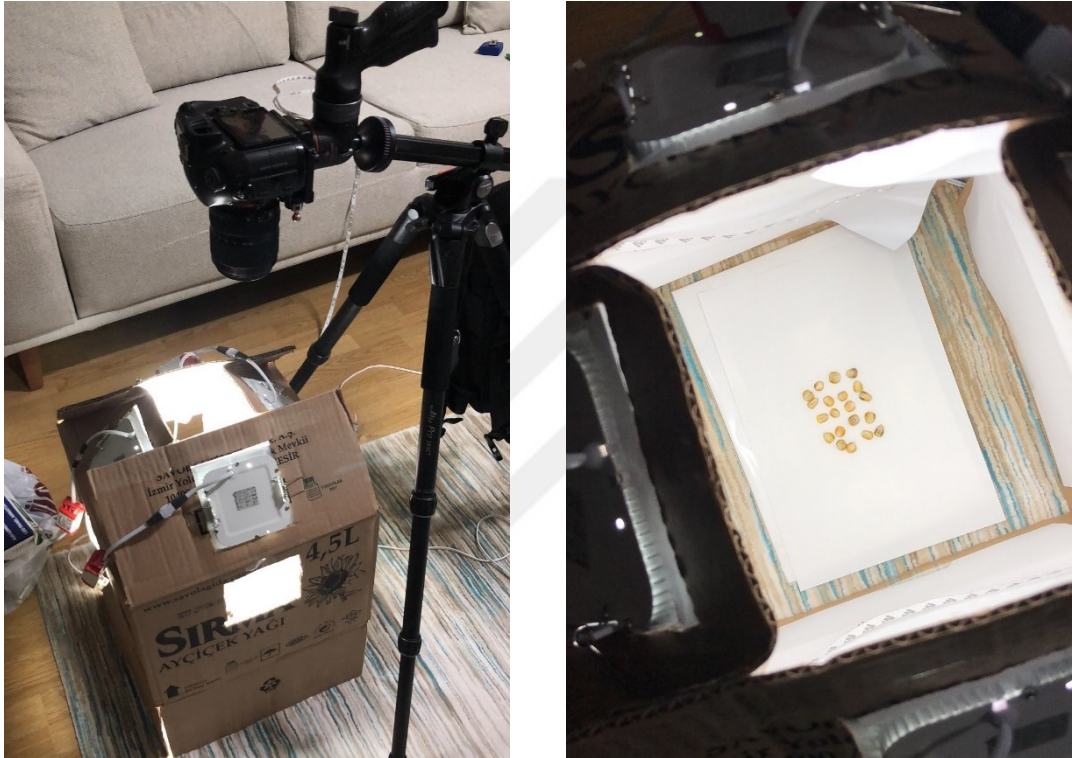
No	Donör	No	Donör	No	Donör	No	Donör
1	TTSM A.Ü.-1 ÇN-1	41	TTSM A.Ü.-2 ÇN-22	81	TTSM SILAJ ÇN-7	121	TTSM II. ÜRÜN-2 ÇN-8
2	TTSM A.Ü.-1 ÇN-2	42	SC SIMBA	82	TTSM SILAJ ÇN-8	122	TTSM II. ÜRÜN-2 ÇN-9
3	TTSM A.Ü.-1 ÇN-3	43	BTS 131412	83	TTSM SILAJ ÇN-9	123	TTSM II. ÜRÜN-2 ÇN-10
4	TTSM A.Ü.-1 ÇN-4	44	PAN 14-701	84	TTSM SILAJ ÇN-10	124	TTSM II. ÜRÜN-2 ÇN-11
5	TTSM A.Ü.-1 ÇN-6	45	WARNER 440C	85	TTSM SILAJ ÇN-11	125	TTSM II. ÜRÜN-2 ÇN-12
6	TTSM A.Ü.-1 ÇN-7	46	SHS 7090	86	TTSM SILAJ ÇN-12	126	TTSM II. ÜRÜN-2 ÇN-13
7	TTSM A.Ü.-1 ÇN-8	47	SHS 5050	87	TTSM II. ÜRÜN-1 ÇN-1	127	TTSM II. ÜRÜN-2 ÇN-14
8	TTSM A.Ü.-1 ÇN-9	48	ANADON	88	TTSM II. ÜRÜN-1 ÇN-2	128	TTSM II. ÜRÜN-2 ÇN-15
9	TTSM A.Ü.-1 ÇN-10	49	SEITEC 6633	89	TTSM II. ÜRÜN-1 ÇN-3	129	TTSM II. ÜRÜN-2 ÇN-16
10	TTSM A.Ü.-1 ÇN-11	50	BM 3063	90	TTSM II. ÜRÜN-1 ÇN-4	130	TTSM II. ÜRÜN-2 ÇN-17
11	TTSM A.Ü.-1 ÇN-12	51	SC TEMBO 73	91	TTSM II. ÜRÜN-1 ÇN-5	131	TTSM II. ÜRÜN-2 ÇN-18
12	TTSM A.Ü.-1 ÇN-13	52	W440C	92	TTSM II. ÜRÜN-1 ÇN-6	132	TTSM II. ÜRÜN-2 ÇN-19
13	TTSM A.Ü.-1 ÇN-14	53	W200C	93	TTSM II. ÜRÜN-1 ÇN-7	133	TTSM II. ÜRÜN-2 ÇN-20
14	TTSM A.Ü.-1 ÇN-15	54	SHS 5090	94	TTSM II. ÜRÜN-1 ÇN-8	134	TTSM II. ÜRÜN-2 ÇN-21
15	TTSM A.Ü.-1 ÇN-16	55	SEITEC 6221	95	TTSM II. ÜRÜN-1 ÇN-9	135	TTSM II. ÜRÜN-2 ÇN-22
16	TTSM A.Ü.-1 ÇN-17	56	STINE 9734	96	TTSM II. ÜRÜN-1 ÇN-10	136	TTSM II. ÜRÜN-2 ÇN-23
17	TTSM A.Ü.-1 ÇN-18	57	STINE 9633	97	TTSM II. ÜRÜN-1 ÇN-11	137	TTSM II. ÜRÜN-2 ÇN-24
18	TTSM A.Ü.-1 ÇN-19	58	SC DIMA 43	98	TTSM II. ÜRÜN-1 ÇN-12	138	TTSM II. ÜRÜN-2 ÇN-25
19	TTSM A.Ü.-1 ÇN-20	59	STINE EX 109-16-7	99	TTSM II. ÜRÜN-1 ÇN-13	139	TTSM II. ÜRÜN-2 ÇN-26
20	TTSM A.Ü.-1 ÇN-21	60	STINE 9537	100	TTSM II. ÜRÜN-1 ÇN-14	140	TTSM II. ÜRÜN-2 ÇN-27
21	TTSM A.Ü.-2 ÇN-1	61	WARNER W220C	101	TTSM II. ÜRÜN-1 ÇN-15	141	TTSM II. ÜRÜN-2 ÇN-28
22	TTSM A.Ü.-2 ÇN-2	62	STINE EX 110-16-5	102	TTSM II. ÜRÜN-1 ÇN-16	142	ANT-QPM-13
23	TTSM A.Ü.-2 ÇN-3	63	6W29	103	TTSM II. ÜRÜN-1 ÇN-17	143	ANT-QPM-14
24	TTSM A.Ü.-2 ÇN-4	64	XIANGNUO 6.9	104	TTSM II. ÜRÜN-1 ÇN-18	144	CORN BORER COMPOSITE #1
25	TTSM A.Ü.-2 ÇN-6	65	JINNUO 628	105	TTSM II. ÜRÜN-1 ÇN-19	145	CORN BORER COMPOSITE #2
26	TTSM A.Ü.-2 ÇN-7	66	W706	106	TTSM II. ÜRÜN-1 ÇN-20	146	CORN BORER COMPOSITE #3
27	TTSM A.Ü.-2 ÇN-8	67	701	107	TTSM II. ÜRÜN-1 ÇN-21	147	BSCB7
28	TTSM A.Ü.-2 ÇN-9	68	JINNONG KONUAL	108	TTSM II. ÜRÜN-1 ÇN-22	148	BSCB8
29	TTSM A.Ü.-2 ÇN-10	69	6W799	109	TTSM II. ÜRÜN-1 ÇN-23	149	DKC7211
30	TTSM A.Ü.-2 ÇN-11	70	SPANISH WAXY-1 F2	110	TTSM II. ÜRÜN-1 ÇN-24	150	DKC7221
31	TTSM A.Ü.-2 ÇN-12	71	SPANISH WAXY-2 F2	111	TTSM II. ÜRÜN-1 ÇN-25		
32	TTSM A.Ü.-2 ÇN-13	72	SPANISH RAW WAXY CORN-1	112	TTSM II. ÜRÜN-1 ÇN-26		
33	TTSM A.Ü.-2 ÇN-14	73	SPANISH RAW WAXY CORN-2	113	TTSM II. ÜRÜN-1 ÇN-27		
34	TTSM A.Ü.-2 ÇN-15	74	WAXY F2	114	TTSM II. ÜRÜN-2 ÇN-1		
35	TTSM A.Ü.-2 ÇN-16	75	TTSM SILAJ ÇN-1	115	TTSM II. ÜRÜN-2 ÇN-2		
36	TTSM A.Ü.-2 ÇN-17	76	TTSM SILAJ ÇN-2	116	TTSM II. ÜRÜN-2 ÇN-3		
37	TTSM A.Ü.-2 ÇN-18	77	TTSM SILAJ ÇN-3	117	TTSM II. ÜRÜN-2 ÇN-4		
38	TTSM A.Ü.-2 ÇN-19	78	TTSM SILAJ ÇN-4	118	TTSM II. ÜRÜN-2 ÇN-5		
39	TTSM A.Ü.-2 ÇN-20	79	TTSM SILAJ ÇN-5	119	TTSM II. ÜRÜN-2 ÇN-6		
40	TTSM A.Ü.-2 ÇN-21	80	TTSM SILAJ ÇN-6	120	TTSM II. ÜRÜN-2 ÇN-7		

Çizelge 3.2 KH2017 verisetinde kullanılan donör genotipler

No	Donör	No	Donör	No	Donör
1	TTSM A.Ü.1 ÇN-1	41	TTSM A.Ü.2 ÇN-21	81	TTSM SİLAJ ÇN-3
2	TTSM A.Ü.1 ÇN-2	42	TTSM A.Ü.2 ÇN-22	82	TTSM SİLAJ ÇN-4
3	TTSM A.Ü.1 ÇN-3	43	TTSM II.Ü.1 ÇN-1	83	TTSM SİLAJ ÇN-5
4	TTSM A.Ü.1 ÇN-4	44	TTSM II.Ü.1 ÇN-2	84	TTSM SİLAJ ÇN-6
5	TTSM A.Ü.1 ÇN-5	45	TTSM II.Ü.1 ÇN-3	85	TTSM SİLAJ ÇN-7
6	TTSM A.Ü.1 ÇN-6	46	TTSM II.Ü.1 ÇN-4	86	TTSM SİLAJ ÇN-8
7	TTSM A.Ü.1 ÇN-7	47	TTSM II.Ü.1 ÇN-5	87	TTSM SİLAJ ÇN-9
8	TTSM A.Ü.1 ÇN-8	48	TTSM II.Ü.1 ÇN-6	88	TTSM SİLAJ ÇN-11
9	TTSM A.Ü.1 ÇN-9	49	TTSM II.Ü.1 ÇN-7	89	TTSM SİLAJ ÇN-12
10	TTSM A.Ü.1 ÇN-10	50	TTSM II.Ü.1 ÇN-9	90	TTSM SİLAJ ÇN-13
11	TTSM A.Ü.1 ÇN-11	51	TTSM II.Ü.1 ÇN-10	91	TTSM SİLAJ ÇN-14
12	TTSM A.Ü.1 ÇN-12	52	TTSM II.Ü.1 ÇN-12	92	TTSM SİLAJ ÇN-15
13	TTSM A.Ü.1 ÇN-14	53	TTSM II.Ü.1 ÇN-13	93	SPANISH RAW WAXY CORN-1
14	TTSM A.Ü.1 ÇN-15	54	TTSM II.Ü.1 ÇN-14	94	SPANISH RAW WAXY CORN-2
15	TTSM A.Ü.1 ÇN-16	55	TTSM II.Ü.2 ÇN-1	95	SPANISH WAXY-1 F2
16	TTSM A.Ü.1 ÇN-17	56	TTSM II.Ü.2 ÇN-2	96	SPANISH WAXY-2 F2
17	TTSM A.Ü.1 ÇN-19	57	TTSM II.Ü.2 ÇN-3	97	WAXY CORN F2
18	TTSM A.Ü.1 ÇN-20	58	TTSM II.Ü.2 ÇN-4	98	P1429
19	TTSM A.Ü.1 ÇN-22	59	TTSM II.Ü.2 ÇN-5	99	P1570
20	TTSM A.Ü.1 ÇN-23	60	TTSM II.Ü.2 ÇN-6	100	P2088
21	TTSM A.Ü.2 ÇN-1	61	TTSM II.Ü.2 ÇN-7	101	CORNELL COLD TOL. SYN-2
22	TTSM A.Ü.2 ÇN-2	62	TTSM II.Ü.2 ÇN-8	102	CORNELL COLD TOL. SYN-3
23	TTSM A.Ü.2 ÇN-3	63	TTSM II.Ü.2 ÇN-9	103	DRK. SPA DROUGHT SYN-1
24	TTSM A.Ü.2 ÇN-4	64	TTSM II.Ü.2 ÇN-10	104	DRK. SPA DROUGHT SYN-2
25	TTSM A.Ü.2 ÇN-5	65	TTSM II.Ü.2 ÇN-11	105	DRK. SPA DROUGHT SYN-3
26	TTSM A.Ü.2 ÇN-6	66	TTSM II.Ü.2 ÇN-12	106	6100-A (WAXY)
27	TTSM A.Ü.2 ÇN-7	67	TTSM II.Ü.2 ÇN-13	107	WAXY F2 İSPANYA
28	TTSM A.Ü.2 ÇN-8	68	TTSM II.Ü.2 ÇN-14		
29	TTSM A.Ü.2 ÇN-9	69	TTSM II.Ü.2 ÇN-15		
30	TTSM A.Ü.2 ÇN-10	70	TTSM II.Ü.2 ÇN-16		
31	TTSM A.Ü.2 ÇN-11	71	TTSM II.Ü.2 ÇN-17		
32	TTSM A.Ü.2 ÇN-12	72	TTSM II.Ü.2 ÇN-18		
33	TTSM A.Ü.2 ÇN-13	73	TTSM II.Ü.2 ÇN-19		
34	TTSM A.Ü.2 ÇN-14	74	TTSM II.Ü.2 ÇN-20		
35	TTSM A.Ü.2 ÇN-15	75	TTSM II.Ü.2 ÇN-21		
36	TTSM A.Ü.2 ÇN-16	76	TTSM II.Ü.2 ÇN-22		
37	TTSM A.Ü.2 ÇN-17	77	TTSM II.Ü.2 ÇN-23		
38	TTSM A.Ü.2 ÇN-18	78	TTSM II.Ü.2 ÇN-24		
39	TTSM A.Ü.2 ÇN-19	79	TTSM SİLAJ ÇN-1		
40	TTSM A.Ü.2 ÇN-20	80	TTSM SİLAJ ÇN-2		

3.1.2. Görüntülerin elde edilmesi

Tüm görüntülerin aynı şartlar altında elde edilmesi için bir görüntüleme düzeneği kurulmuştur. Görüntüleme düzeneğinin tavanına kamera ve ışık kaynağı olarak yeterli sayıda LED montajlanmıştır. Tüm tohumlar embriyoları yukarı bakacak pozisyonda ve birbirlerine temas etmeksizin, her görüntüde 20 tohum (4 satır, 5 sütun) olacak şekilde fotoğraflanmıştır. Görüntüleme düzeneğine ait görüntüler Şekil 3.2’de verilmiştir.



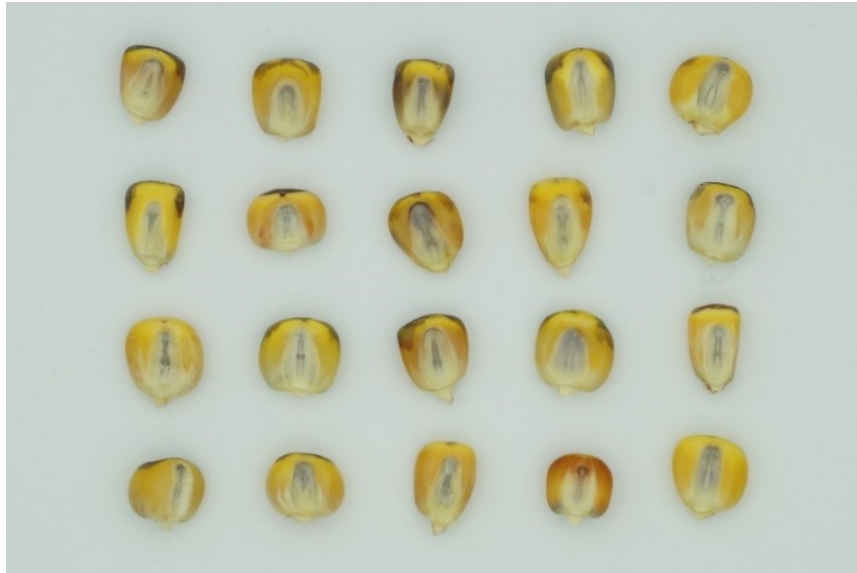
Şekil 3.2 Görüntüleme düzeneği

KH2016 veri setini oluşturan mısır tohumlarının görüntülerin elde edilmesinde Sony ILCE-6000 dijital fotoğraf makinası kullanılmıştır. Tüm görüntüler M modunda, 200 ISO, 1/125 sn enstantane, f/7,1 diyafram açıklığı, 24 mm odak uzaklığı, 18 cm kamera uzaklığında ve 4240 x 2832 piksel çözünürlükte çekilmiştir. Elde edilen ham görüntü örneği Şekil 3.3’de gösterilmiştir.



Şekil 3.3 KH2016 veri seti ham görüntü örneği

KH2017 veri setini oluşturan mısır tohumlarının görüntülerinin elde edilmesinde Sony SLT-A58 dijital fotoğraf makinası kullanılmıştır. Tüm görüntüler M modunda, 100 ISO, 1/125 sn enstantane, f/9 diyafram açıklığı, 150 mm odak uzaklığı, 55 cm kamera uzaklığında ve 5456 x 3632 piksel çözünürlükte çekilmiştir. Elde edilen ham görüntü örneği Şekil 3.4’da gösterilmiştir.



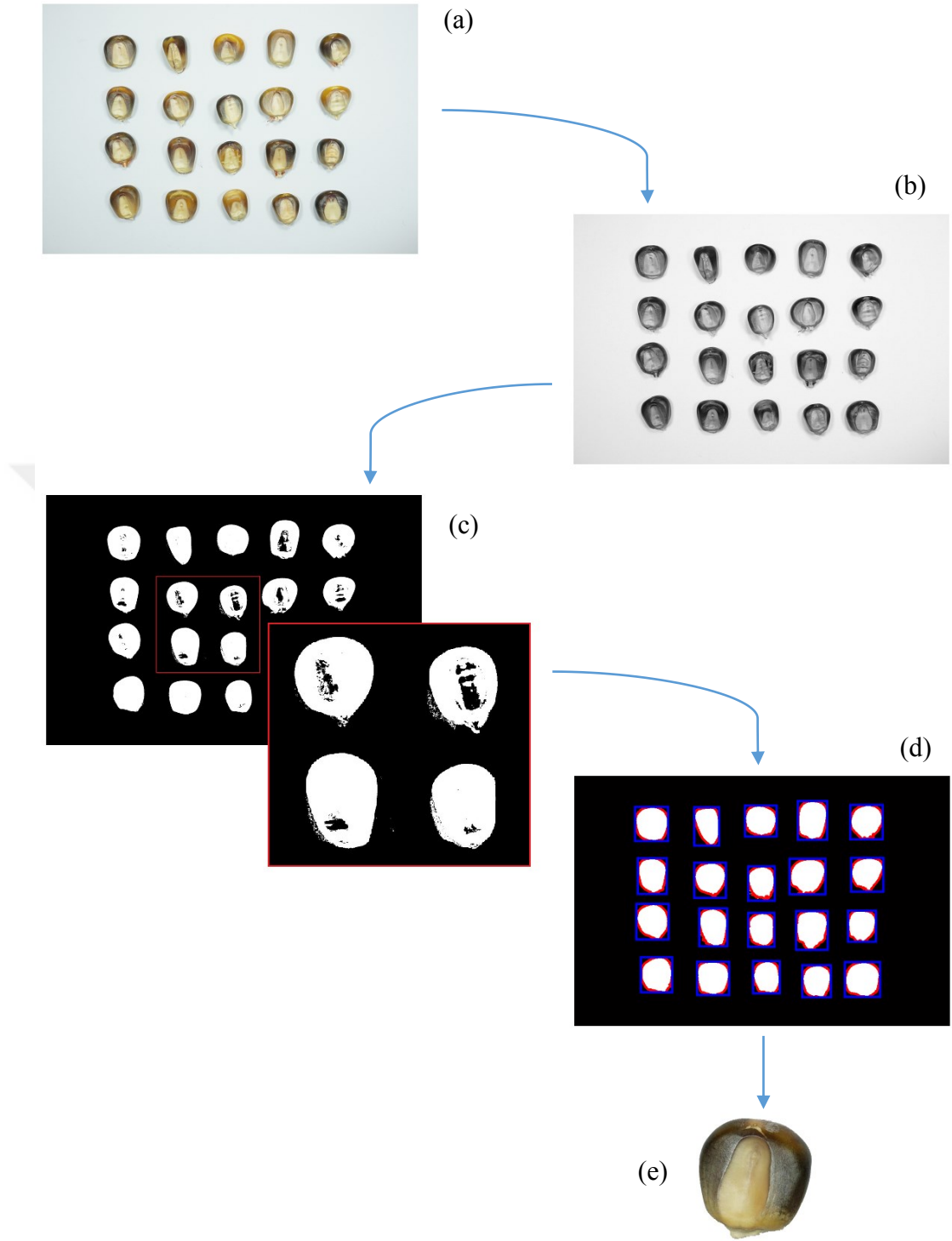
Şekil 3.4 KH2017 veri seti ham görüntü örneği

3.1.3. Mısır tohumlarının imgeden bölütlenmesi

Bölütleme, imge içerisinde ilgi duyulan alanların imgeden ayrılması; başka bir ifadeyle bir imgeyi anlamlı bölgelere ayırma işlemidir. Mısır tohumlarının görüntüden bölütlenmesi işleminde her bir mısır tohumu toplu halde çekilen orijinal görüntüden ayrılmıştır. Görüntü bölütleme konusunda her imgeye uygulanabilecek genel bir yöntem yoktur [24]. Her problem için farklı bir bölütleme yöntemi kullanılır. Mısır tohumlarının imgeden bölütlenmesi işlemi için RGB imgenin B kanalı kullanılmıştır. KH2016 veri seti için 190 eşik değeri, KH2017 veri seti için 150 eşik değeri ampirik olarak belirlenmiştir. Denklem (3.1)'e göre ikili görüntü elde edilmiştir. $B < \text{eşik değeri}$ değeri değerine sahip her piksel mısır tohumu, değilse arka plan olarak belirlenmiştir.

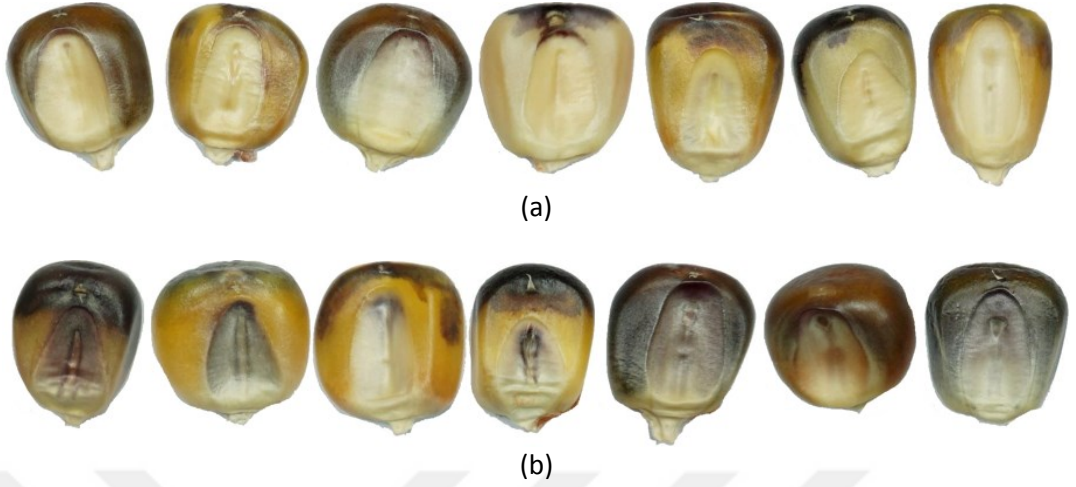
$$g(x, y) = \begin{cases} \text{mısır tohumu,} & B < \text{eşik değeri} \\ \text{arka plan,} & \text{değilse} \end{cases} \quad (3.1)$$

Elde edilen ikili görüntüye aşındırma ve doldurma morfolojik işlemleri uygulanmıştır. Mısır tohumu olarak belirlenen alanların sınır çizgisi kullanılarak her bir tohum için sınırlayıcı kutu hesaplanmış ve arkaplan temizlenmiş olarak ayrı birer imge olarak kaydedilerek bölütleme işlemi tamamlanmıştır. Yapılan işlem adımları Şekil 3.5'de gösterilmiştir.



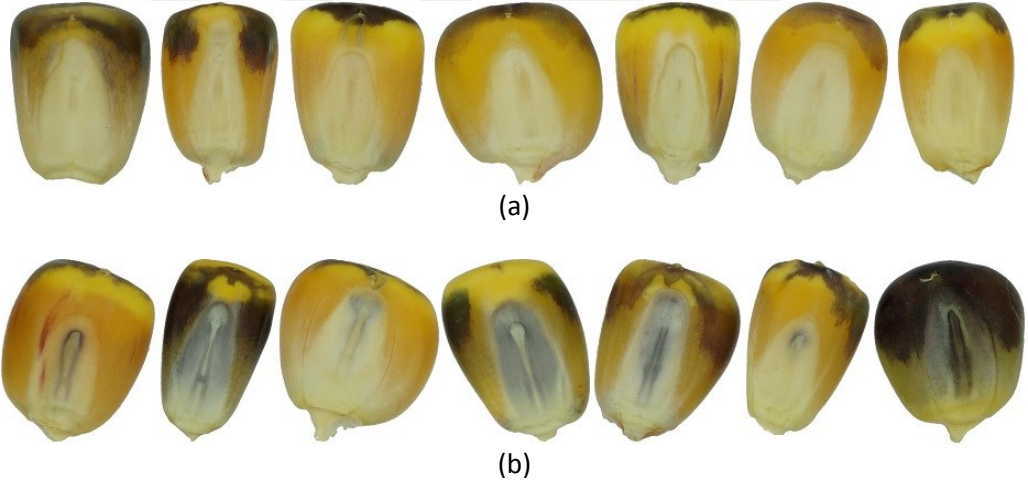
Şekil 3.5 Mısır tohumlarının imgeden bölütlenmesi aşamaları.

KH2016 veri setine ait bölütlenmiş haploid ve diploid mısır tohumu örnekleri Şekil 3.6’da gösterilmiştir.



Şekil 3.6 KH2016 veri setine ait (a) haploid ve (b) diploid mısır tohum örnekleri

KH2017 veri setine ait bölütlenmiş haploid ve diploid mısır tohumu örnekleri Şekil 3.7’da gösterilmiştir.



Şekil 3.7 KH2017 veri setine ait (a) haploid ve (b) diploid mısır tohum örnekleri

3.2. Embriyo Bölütleme

Haploid ve diploid tohumların endospermelerinde R1-nj renklenmesi gözlenmektedir. Bununla birlikte sadece diploidlerin embriyolarında R1-nj renklenmesi oluşurken, baba ebeveynin kromozomlarını taşımadığı için haploidlerin embriyolarında renklenme oluşmamaktadır. Bu sebeple sadece embriyo renklenmelerine bakılarak tohumlar haploid ve diploid olarak sınıflandırılabilir.

In vivo maternal haploid melezleme sonucunda elde edilen hem embriyosunda hem de endospermde renklenme gözlenmeyen tohumlar melez dışı kabul edilmektedir. Bu tohumların kontamine olduğu varsayılır. Sadece embriyo renklenmesine bakılarak yapılacak bir sınıflandırmada melez dışı tohumlarında haploid olarak sınıflandırılacağı açıktır. Bu tez çalışmasında haploid ve diploid mısır tohumlarının sınıflandırılması için sadece tohum embriyosundan özniteliklerin çıkarılmış, melez dışı tohumlar göz ardı edilmiştir. Melez dışı tohumların göz ardı edilmesinin nedeni hem melez dışı tohumların çok düşük bir frekansta olmaları nedeniyle kabul edilebilir bir hataya sebep olacakları, hem de embriyo yukarı pozisyonda renklenme gözlenmeyen ancak gerçekte haploid olan tohumların kaybedilmemesi içindir. R1-nj renklenmesi küçük bir noktadan, tüm endospermi kapsayan renkliliğe kadar çeşitlilik gösterir. Renklenme Şekil 3.8’de gösterildiği gibi imgede görülmeyen bir yerde (tepedeki çukurda veya tohum sırtında) olabilir. Bu tez çalışmasında, söz konusu renklenme gösteren haploid mısır tohumlarının kaybedilmemesi için tüm mısırlar embriyolarına bakılarak, sadece haploid ve diploid olarak sınıflandırılmıştır.



Şekil 3.8 Endospermde R1-nj renklenmesi

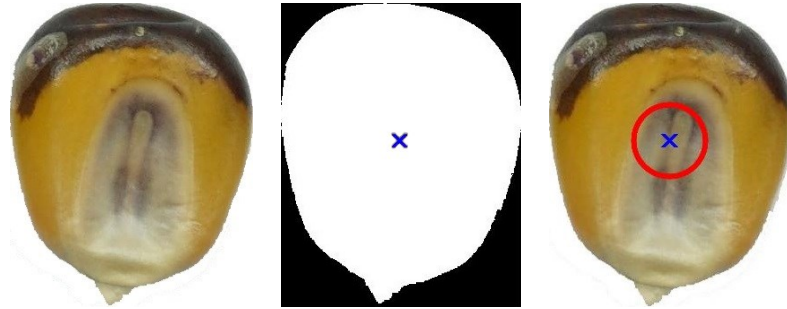
Bu tez çalışmasında söz konusu embriyo bölütleme işlemini gerçekleştirmek için Chu vd. [25] tarafından önerilen yöntem kullanılmıştır. Bu yöntemle göre embriyonun tohum kütle merkezini kapsadığı varsayılır ve tohum kütle merkezli bir çemberin embriyoyu temsil edeceği kabul edilir.

Chu vd. [25] küflenmiş mısır tohumlarını sağlıklı mısır tohumlarından ayırmak için önerdikleri yöntemde merkezi tohumun kütle merkezi ve çapı embriyo genişliği kadar olan bir çember ile embriyo bölgesinin temsil edilebileceğini belirtmişlerdir. Embriyo ve endosperm bölgelerinin tespiti için bu yöntem kullanılmıştır. Denklem (3.1) kullanılarak oluşturulan ikili görüntüden, Denklem (3.2) kullanılarak tohum kütle merkezi hesaplanmıştır.

$$y_0 = \frac{\sum yg(x,y)}{\sum g(x,y)}$$

$$x_0 = \frac{\sum xg(x,y)}{\sum g(x,y)}$$
(3.2)

KH2016 veri seti içerisinde rastgele seçilen 80 mısır tohumunda yapılan ölçümlerde istatistiksel olarak 35 piksel yarıçaplı çemberin embriyoyu iyi temsil ettiği tespit edilmiştir. KH2017 veri seti içerisinde rastgele seçilen 80 mısır tohumunda yapılan ölçümlerde istatistiksel olarak 50 piksel yarıçaplı çemberin embriyoyu iyi temsil ettiği tespit edilmiştir. Hesaplanan çemberin içi embriyo, dışı endosperm olarak kabul edilmiştir. Şekil 3.9’de diploid mısır tohum örneğinde tohum kütle merkezi ve öznitelik çıkarımı yapılan bölge gösterilmiştir.



Şekil 3.9 Tohum kütle merkezi ve öznitelik çıkarımı yapılan bölge

3.3. Öznitelik Çıkarımı

Eğitmenli makine öğrenmesinde sınıflandırma yapabilmek için ham veriden (imge) sınıflandırıcılar tarafından kullanılacak anlamsal sayısal değerlerin türetilmesi işlemine öznitelik çıkarımı denilmektedir. Makine öğrenimi alanında kullanılan

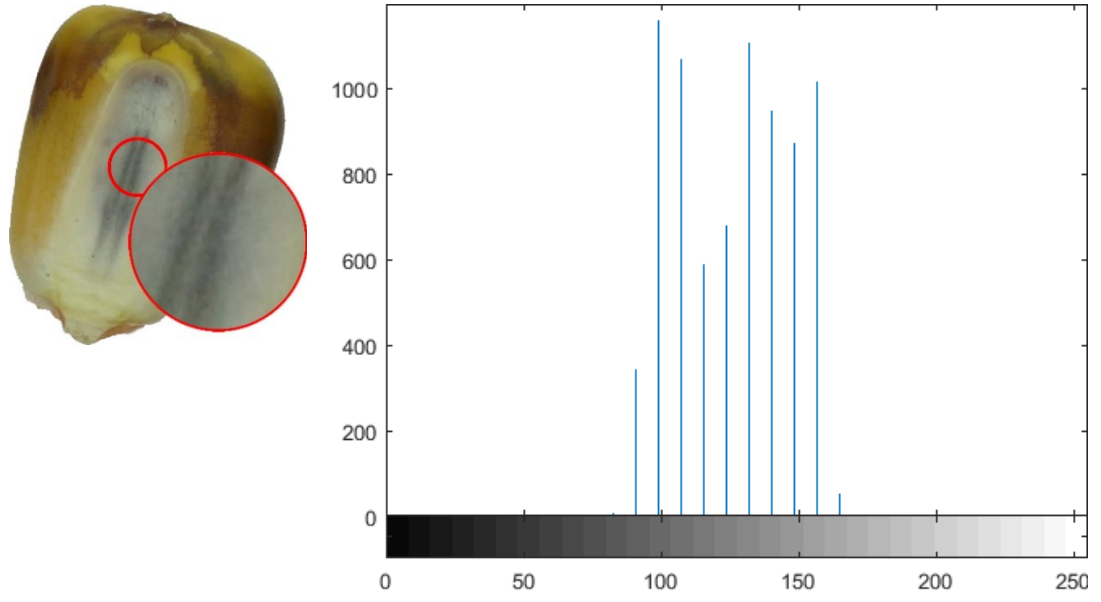
öznitelik çıkarımı, girdi olarak verilen ölçülmüş verileri kullanarak türetilmiş değerler oluşturur. Özellik çıkarımı işlemi bir boyut azaltma işlemidir. Öznitelik çıkarımı sonucunda elde edilen özniteliklerin karşılığını tutan veri yapısına öznitelik vektörü denilmektedir. Öznitelik vektörleri kullanılarak sınıflandırıcılar eğitilir ve sınıf etiketini bilmedikleri örnekler için başarılı bir sınıf tahmini yapmaları beklenir.

Bu tez çalışmasında, her bir veri seti için beş farklı öznitelik vektörü çıkarılmıştır. Bölütlenen embriyo bölgelerinden renk ve doku öznitelikleri çıkarılmıştır. Renk öznitelikleri olarak renk histogramları ve renk momentleri, doku özniteliği olarak gri seviyeli eş oluşum matrisi kullanılmıştır. Renk histogram öznitelik vektörleri 8, 16 ve 32 renk ölçekli gri seviye embriyo imgelerinin renk histogram değerleri ayrı ayrı öznitelik vektörü olarak kullanılmıştır. Renk momentleri öznitelik vektörü RGB renk uzayında embriyo imgelerinin her renk kanalından ayrı ayrı olmak üzere ilk dört derece renk momentleri *ortalama*, *standart sapma*, *çarpıklık* ve *basıklık* kullanılarak oluşturulmuştur. Doku öznitelik vektörü embriyo imgelerinin gri seviye eş oluşum matrisinden elde edilen *enerji*, *homojenlik*, *kontrast*, *korelasyon* ve *entropi* değerlerinden oluşturulmuştur.

3.3.1. Renk histogramları

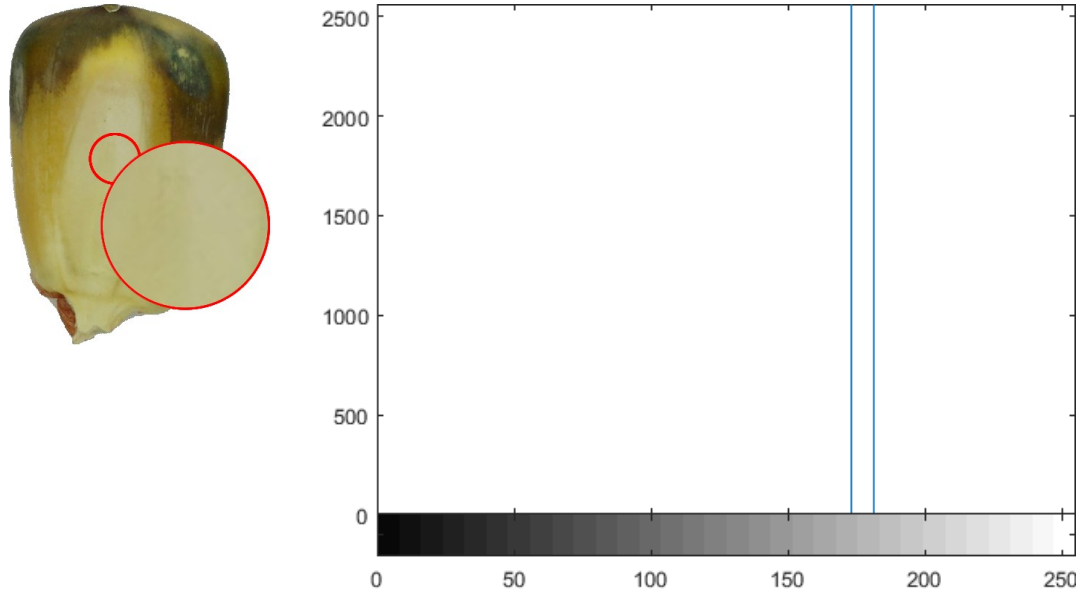
Renk histogramları bir imgede renk öznitelikleri çıkarmak için yaygın olarak kullanılan bir tekniktir [26, 27]. Renk histogramları kolay hesaplanabilir, döndürme ve çevirme gibi imgedeki değişikliklerden etkilenmez ve imgenin renk dağılımını etkili bir şekilde temsil eder [28–30].

Bu tez çalışmasında bölütlenen embriyo bölgesi 8, 16 ve 32 renk ölçekli gri seviyeli imgeye dönüştürülmüş ve elde edilen renk histogramları öznitelik vektörü olarak kullanılmıştır. Şekil 3.10'da örnek bir diploid mısır tohumundan bölütlenen embriyo bölgesinin 32 renk ölçekli gri seviye renk histogramı gösterilmiştir.



Şekil 3.10 Diploid tohum örneğinde öznelilik çıkartılan bölge ve renk histogramı

Şekil 3.11’de haploid mısır tohum örneğinde bölütlenen embriyo bölgesinin 32 renk ölçekli gri seviye renk histogramı gösterilmiştir.



Şekil 3.11 Haploid tohum örneğinde öznelilik çıkartılan bölge ve renk histogramı

Çizelge 3.3 ve Çizelge 3.4’de KH2016 ve KH2017 veri setleri için 8 renk ölçekli gri seviye renk histogram özniteliklerinin haploid ve diploid sınıflarına göre aldıkları *en küçük, en büyük, ortalama ve standart sapma* değerleri verilmiştir.

Çizelge 3.3 KH2016 veri seti 8 renk ölçekli gri seviye renk histogram özniteliklerinin aldığı değerler

Öznitelikler	Haploid sınıfı				Diploid sınıfı			
	EK	EB	Ort	SD	EK	EB	Ort	SD
Renk1	0	0	0	0	0	561	3,29	35,82
Renk2	0	158	1,82	16,94	0	2685	63,61	260,12
Renk3	0	264	3,70	28,91	0	2617	215,49	476,37
Renk4	0	522	11,18	64,76	0	2710	600,07	693,93
Renk5	0	1562	206,44	361,62	0	2740	930,90	731,25
Renk6	78	2821	2005,32	742,03	0	2819	908,69	938,14
Renk7	0	2743	592,54	784,63	0	2261	98,95	301,26
Renk8	0	0	0	0	0	0	0	0

EK: En küçük değer, EB: En büyük değer, Ort: Ortalama, SD: Standart sapma

Çizelge 3.4 KH2017 veri seti 8 renk ölçekli gri seviye renk histogram özniteliklerinin aldığı değerler

Öznitelikler	Haploid sınıfı				Diploid sınıfı			
	EK	EB	Ort	SD	EK	EB	Ort	SD
Renk1	0	0	0	0	0	63	0,04	1,50
Renk2	0	10	0,01	0,29	0	2116	7,23	88,67
Renk3	0	1594	4,29	76,84	0	6562	189,08	721,96
Renk4	0	7613	51,54	410,23	0	7845	1242,05	1848,24
Renk5	0	7845	2039,47	2535,22	0	7845	4198,99	2211,86
Renk6	0	7845	5749,69	2606,82	0	7845	2207,62	2393,01
Renk7	0	0	0	0	0	0	0	0
Renk8	0	0	0	0	0	0	0	0

EK: En küçük değer, EB: En büyük değer, Ort: Ortalama, SD: Standart sapma

Çizelge 3.5 ve Çizelge 3.6’da KH2016 ve KH2017 veri setleri için 16 renk ölçekli gri seviye renk histogram özniteliklerinin haploid ve diploid sınıflarına göre aldıkları *en küçük, en büyük, ortalama ve standart sapma* değerleri verilmiştir.

Çizelge 3.5 KH2016 veri seti 16 renk ölçekli gri seviye renk histogram
özniteliklerinin aldığı değerler

Öznitelikler	Haploid sınıfı				Diploid sınıfı			
	EK	EB	Ort	SD	EK	EB	Ort	SD
Renk1	0	0	0	0	0	2	0,01	0,11
Renk2	0	0	0	0	0	1077	10,52	92,85
Renk3	0	12	0,14	1,29	0	1554	25,57	139,38
Renk4	0	183	2,10	19,62	0	1464	51,62	171,28
Renk5	0	112	1,30	12,01	0	2216	89,71	250,58
Renk6	0	126	2,26	15,43	0	1286	123,60	234,19
Renk7	0	173	3,53	23,33	0	1745	230,65	337,64
Renk8	0	335	7,37	40,52	0	2047	350,79	417,72
Renk9	0	558	27,21	101,23	0	1696	421,30	419,75
Renk10	0	1220	134,97	248,19	0	1929	452,32	438,03
Renk11	0	2508	681,91	675,11	0	2259	510,69	547,83
Renk12	3	2619	1093,99	676,46	0	2419	384,98	545,31
Renk13	0	2584	746,48	753,44	0	2216	151,46	350,36
Renk14	0	1972	119,62	289,84	0	886	17,79	88,91
Renk15	0	10	0,13	1,08	0	1	0,01	0,08
Renk16	0	0	0	0	0	0	0	0

EK: En küçük değer, EB: En büyük değer, Ort: Ortalama, SD: Standart sapma

Çizelge 3.6 KH2017 veri seti 16 renk ölçekli gri seviye renk histogram
özniteliklerinin aldığı değerler

Öznitelikler	Haploid sınıfı				Diploid sınıfı			
	EK	EB	Ort	SD	EK	EB	Ort	SD
Renk1	0	0	0	0	0	0	0	0
Renk2	0	0	0	0	0	227	0,13	5,40
Renk3	0	0	0	0	0	531	0,88	16,91
Renk4	0	42	0,04	1,25	0	2217	12,16	115,50
Renk5	0	527	0,85	17,78	0	3199	50,32	267,69
Renk6	0	1704	3,96	70,11	0	4559	160,60	543,91
Renk7	0	2966	11,55	147,78	0	6753	378,03	831,14
Renk8	0	4869	39,45	289,95	0	7141	836,26	1199,16
Renk9	0	7671	291,36	899,65	0	7221	1534,78	1448,30
Renk10	0	7837	1404,78	1920,05	0	7107	2341,50	1623,13
Renk11	0	7845	4408,32	2392,06	0	7834	2186,55	2088,42
Renk12	0	7845	1683,48	2286,82	0	6458	342,28	938,00
Renk13	0	1173	1,22	34,44	0	2619	1,50	62,26
Renk14	0	0	0	0	0	0	0	0
Renk15	0	0	0	0	0	0	0	0
Renk16	0	0	0	0	0	0	0	0

EK: En küçük değer, EB: En büyük değer, Ort: Ortalama, SD: Standart sapma

Çizelge 3.7 ve Çizelge 3.8’de KH2016 ve KH2017 veri setleri için 32 renk ölçekli gri seviye renk histogram özniteliklerinin haploid ve diploid sınıflarına göre aldıkları *en küçük, en büyük, ortalama ve standart sapma* değerleri verilmiştir.

Çizelge 3.7 KH2016 veri seti 32 renk ölçekli gri seviye renk histogram özniteliklerinin aldığı değerler

Öznitelikler	Haploid sınıfı				Diploid sınıfı			
	EK	EB	Ort	SD	EK	EB	Ort	SD
Renk1	0	0	0	0	0	1	0,01	0,06
Renk2	0	0	0	0	0	92	0,40	5,40
Renk3	0	0	0	0	0	598	4,95	46,54
Renk4	0	0	0	0	0	1013	8,54	70,59
Renk5	0	0	0	0	0	968	13,08	79,40
Renk6	0	45	0,52	4,82	0	526	15,00	64,83
Renk7	0	96	1,10	10,29	0	454	21,15	71,07
Renk8	0	69	0,79	7,40	0	1339	33,89	119,92
Renk9	0	54	0,62	5,79	0	1374	40,46	126,61
Renk10	0	55	0,66	5,90	0	1079	52,00	131,20
Renk11	0	64	1,08	7,54	0	748	53,49	115,54
Renk12	0	59	1,20	7,91	0	657	67,76	114,52
Renk13	0	54	1,13	7,43	0	930	95,40	153,33
Renk14	0	142	2,63	17,76	0	1066	142,24	201,55
Renk15	0	124	2,31	15,45	0	1126	156,41	205,31
Renk16	0	193	4,67	23,97	0	895	177,66	211,62
Renk17	0	249	10,30	45,59	0	1081	195,17	218,08
Renk18	0	284	14,60	51,82	0	1078	200,70	211,48
Renk19	0	483	34,83	89,96	0	1212	230,55	241,34
Renk20	0	761	82,71	148,93	0	1437	219,20	237,63
Renk21	0	1222	229,94	305,08	0	1650	236,34	271,96
Renk22	0	1721	356,32	364,08	0	1565	245,16	302,28
Renk23	0	1893	564,37	424,59	0	1788	241,39	334,04
Renk24	0	1793	528,40	430,30	0	1479	167,68	276,30
Renk25	0	1662	443,66	379,32	0	1351	114,78	234,24
Renk26	0	1410	332,83	419,10	0	1315	57,98	163,81
Renk27	0	1236	173,69	316,98	0	1107	26,13	122,65
Renk28	0	1200	32,34	136,75	0	270	3,44	22,24
Renk29	0	20	0,29	2,16	0	2	0,02	0,17
Renk30	0	2	0,02	0,21	0	1	0,01	0,06
Renk31	0	0	0	0	0	0	0	0
Renk32	0	0	0	0	0	0	0	0

EK: En küçük değer, EB: En büyük değer, Ort: Ortalama, SD: Standart sapma

Çizelge 3.8 KH2017 veri seti 32 renk ölçekli gri seviye renk histogram
özniteliklerinin aldığı değerler

Öznitelikler	Haploid sınıfı				Diploid sınıfı			
	EK	EB	Ort	SD	EK	EB	Ort	SD
Renk1	0	0	0	0	0	0	0	0
Renk2	0	0	0	0	0	0	0	0
Renk3	0	0	0	0	0	116	0,07	2,76
Renk4	0	0	0	0	0	168	0,10	3,99
Renk5	0	0	0	0	0	171	0,24	5,31
Renk6	0	0	0	0	0	832	1,55	27,94
Renk7	0	9	0,01	0,26	0	1108	4,58	54,82
Renk8	0	66	0,06	1,89	0	1270	10,09	79,62
Renk9	0	227	0,23	6,61	0	1850	19,05	119,09
Renk10	0	542	1,01	19,89	0	1739	39,11	182,79
Renk11	0	794	1,85	35,36	0	2302	64,32	245,37
Renk12	0	959	1,97	35,56	0	2810	100,33	314,36
Renk13	0	1167	3,00	44,97	0	4248	149,57	383,79
Renk14	0	2873	9,85	127,07	0	5494	246,75	513,96
Renk15	0	3071	15,27	144,20	0	3760	340,26	570,84
Renk16	0	2465	22,59	152,33	0	5214	462,38	656,85
Renk17	0	4321	66,14	311,83	0	5594	634,68	737,63
Renk18	0	6822	189,42	611,89	0	5611	789,93	780,03
Renk19	0	6099	433,22	875,25	0	6030	1090,49	891,63
Renk20	0	7315	853,12	1221,73	0	5679	1201,12	998,81
Renk21	0	7319	1631,10	1483,34	0	5700	1279,67	1182,94
Renk22	0	7520	2364,65	1788,48	0	6146	919,88	1159,11
Renk23	0	7826	1977,99	2185,92	0	5999	436,86	1007,72
Renk24	0	6615	269,18	818,51	0	3708	50,85	281,64
Renk25	0	3042	4,34	94,49	0	4909	3,13	117,10
Renk26	0	0	0	0	0	0	0	0
Renk27	0	0	0	0	0	0	0	0
Renk28	0	0	0	0	0	0	0	0
Renk29	0	0	0	0	0	0	0	0
Renk30	0	0	0	0	0	0	0	0
Renk31	0	0	0	0	0	0	0	0
Renk32	0	0	0	0	0	0	0	0

EK: En küçük değer, EB: En büyük değer, Ort: Ortalama, SD: Standart sapma

3.3.2. Renk momentleri

Renk insanların bir görüntüye bakarken algıladığı önemli ve en doğru öz niteliktir. İnsan gözü renkli görüntüye, gri seviyeli görüntüden daha duyarlı olması rengin önemli bir öz nitelik olmasına neden olmaktadır [31]. Sınıflandırma problemlerinde en önemli işlerden biri sınıflandırmaya esas öz nitelik vektörlerinin çıkarımıdır. Renkli imge sınıflandırmada renk öz nitelikleri en yaygın kullanılan görsel öz niteliklerdir [32]. Renk momentleri renkli bir imgedeki renk dağılımının karakteristiğini temsil eden öz niteliklerdir.

Bu tez çalışmasında, RGB renk uzayında her renk kanalı için ayrı ayrı olmak üzere ilk dört dereceden renk momentleri öz nitelik olarak kullanılmıştır. Birinci derece renk momenti *ortalama*yı, ikinci derece renk momenti *standart sapma*yı, üçüncü derece renk momenti *çarpıklık*ı, dördüncü derece renk momenti ise *basıklık*ı ifade etmektedir. İlk dört dereceden renk momentlerinin denklemleri aşağıda verilmiştir.

Ortalama, piksel yoğunluk değerlerinin toplamı piksel sayısına bölünmesiyle elde edilir. Aşağıdaki formülle hesaplanır.

$$\mu_i = \frac{1}{N} \sum_{j=1}^N f_{ij} \quad (3.3)$$

Standart sapma, piksel yoğunluk değerlerinin ortalama yoğunluk değerine olan uzaklıkları toplamını ifade eder. Aşağıdaki formülle hesaplanır.

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^2} \quad (3.4)$$

Çarpıklık, piksel yoğunluk değerlerinin ortalamaya göre simetrik olmama miktarını ölçer. Aşağıdaki formülle hesaplanır.

$$s_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^3} \quad (3.5)$$

Basıklık, Kurtosis olarak da bilinir. Piksel yoğunluk değerleri dağılımının ne kadar dik veya ne kadar basık olduğunu ifade eder. Aşağıdaki formülle hesaplanır.

$$k_i = \sqrt[4]{\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^4} \quad (3.6)$$

Çizelge 3.9 ve Çizelge 3.10’da KH2016 ve KH2017 veri setleri için renk moment özniteliklerinin haploid ve diploid sınıflarına göre aldıkları *en küçük, en büyük, ortalama* ve *standart sapma* değerleri verilmiştir.

Çizelge 3.9 KH2016 veri seti renk momentleri özniteliklerinin aldığı değerler

Öznitelikler		Haploid sınıfı				Diploid sınıfı			
		EK	EB	Ort	SD	EK	EB	Ort	SD
R	Ortalama	165,91	220,92	199,80	10,72	33,21	215,32	157,91	33,99
	Standart sapma	3,56	43,61	8,05	5,19	3,83	51,11	15,75	7,40
	Çarpıklık	-1,72	1,19	-0,09	0,62	-2,50	1,63	-0,28	0,58
	Basıklık	1,67	7,96	3,22	1,13	1,43	10,70	3,11	1,16
G	Ortalama	151,31	216,42	187,98	13,64	29,24	208,26	143,56	35,43
	Standart sapma	4,24	46,40	10,44	5,78	4,53	51,57	17,69	7,26
	Çarpıklık	-1,58	1,20	-0,16	0,55	-2,58	1,74	-0,23	0,59
	Basıklık	1,49	7,52	3,02	0,97	1,38	11,30	3,04	1,06
B	Ortalama	102,56	184,48	140,77	20,38	15,70	174,91	108,00	30,60
	Standart sapma	5,09	36,49	14,88	5,88	4,46	42,58	17,72	6,42
	Çarpıklık	-1,44	1,34	0,01	0,59	-2,09	2,26	-0,01	0,64
	Basıklık	1,44	5,39	2,92	0,91	1,46	11,21	3,10	1,21

EK: En küçük değer, EB: En büyük değer, Ort: Ortalama, SD: Standart sapma

Çizelge 3.10 KH2017 veri seti renk momentleri özniteliklerinin aldığı değerler

Öznitelikler		Haploid sınıfı				Diploid sınıfı			
		EK	EB	Ort	SD	EK	EB	Ort	SD
R	Ortalama	133,25	199,14	183,49	6,73	74,87	197,53	157,48	18,59
	Standart sapma	1,32	39,11	4,12	2,16	1,85	36,74	11,57	4,83
	Çarpıklık	-6,74	1,60	-0,52	0,67	-3,49	1,40	-0,41	0,53
	Basıklık	1,39	81,84	3,82	4,41	1,31	22,29	2,83	1,48
G	Ortalama	103,67	194,39	172,76	11,51	73,91	196,80	150,41	21,52
	Standart sapma	1,18	43,55	6,18	3,28	2,26	43,08	12,13	4,74
	Çarpıklık	-7,18	1,61	-0,59	0,67	-3,81	1,83	-0,43	0,53
	Basıklık	1,27	87,93	3,62	3,66	1,19	20,98	2,85	1,44
B	Ortalama	24,25	148,93	111,00	19,66	10,63	158,70	107,73	25,13
	Standart sapma	2,20	37,88	10,80	5,78	2,90	40,00	11,34	5,08
	Çarpıklık	-4,23	1,75	-0,46	0,60	-3,79	1,98	-0,31	0,57
	Basıklık	1,30	39,73	3,17	1,87	1,23	18,13	2,92	1,25

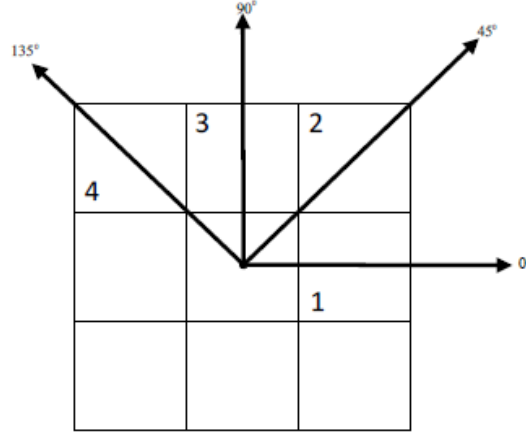
EK: En küçük değer, EB: En büyük değer, Ort: Ortalama, SD: Standart sapma

3.3.3. Gri seviyeli eş oluşum matrisi

Haralick vd. [33] tarafından önerilen gri seviyeli eş oluşum matrisi (GSEM) imgelerden doku özniteliklerin çıkarılmasında kullanılmaktadır. Belirli bir mesafe ve açıdaki iki pikselin değerlerinin imgede görülme sıklığı GSEM'ini vermektedir. Bu piksellerden biri referans; diğeri ise komşu piksel olarak ifade edilir [33, 34]. Gri seviyeli görüntünün büyüklüğü $(m \times n)$ ve mesafe vektörü $d = (\Delta x, \Delta y)$ olduğu durumda normalize olmayan GSEM değerleri Denklem (3.7)'deki eşitlikle tanımlanır.

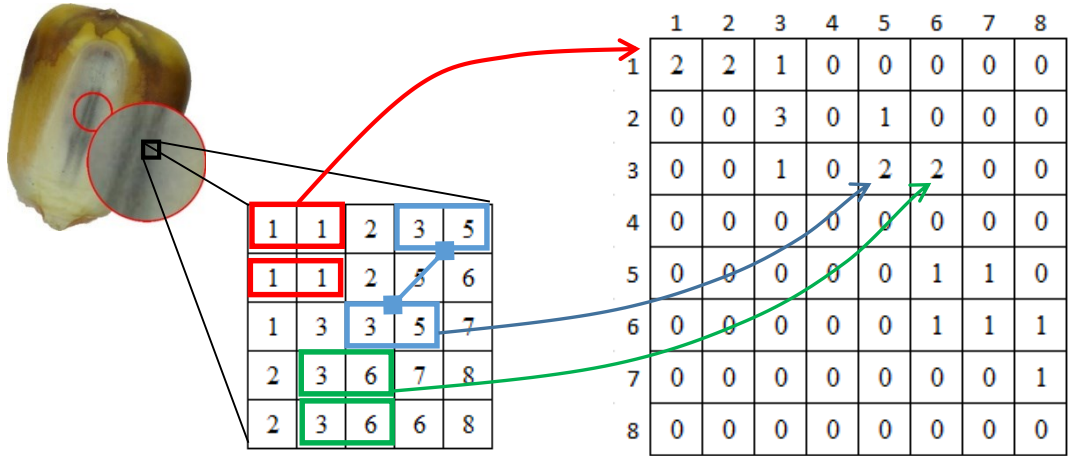
$$P(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1, & I(p, q) = i \text{ ve } I(p + \Delta x, q + \Delta y) = j \text{ ise} \\ 0, & \text{diğer durumlarda} \end{cases} \quad (3.7)$$

GSEM'nin hesaplanmasında ilk aşama renkli imgenin gri seviyeli imgeye dönüştürülmesidir. Gri seviyeli imge üzerinde ön tanımlı mesafe ve açıdaki piksel değerlerinin imgede görülme sıklığı ile GSEM oluşturulur. Pikseller arası uzaklıklar ile birlikte, piksel çiftlerinin yönleri Şekil 3.12'de gösterildiği gibi genellikle $\theta = 0, 45, 90, 135$ olarak kullanılmaktadır.



Şekil 3.12 Referans ve komşu pikseller arasındaki açısal ilişki

Örnek olarak $\delta = 1$, $\theta = 0$ için (5x5) boyutlu ve gri seviye sayısı 8 olan bir imge için GSEM'in oluşturulması Şekil 3.13'de gösterilmiştir. Şekilde gösterildiği gibi, (1,1) piksel çifti imgede 2 kez tekrar ettiği için GSEM matrisinin (1,1) indeksine 2 değeri yazılmıştır. Benzer şekilde (3,5) ve (3,6) piksel çiftleri de imgede 2'şer kez tekrar ettiği için GSEM matrisinde ilgili indekslere 2 değeri yazılmıştır. Bu işlem tüm piksel komşulukları için tekrarlanarak GSEM elde edilir.



Şekil 3.13 GSEM oluşturulması

GSEM matrisinin normalize edilmesinden sonra farklı öznitelikler çıkarılabilmektedir [35]. Bu tez çalışmasında aşağıda tanımları ve eşitlikleri verilen enerji, homojenlik, kontrast, korelasyon ve entropi öznitelikleri kullanılmıştır.

Enerji, imgenin dokusal tekdüzeliğini ölçer. GSEM'in elemanları kareleri toplamı ile hesaplanır ve $[0,1]$ aralığında değerler alır. Tekdüze bir imge için 1 değerini alır ve aşağıdaki eşitlikte gösterildiği gibi hesaplanır.

$$Enerji = \sum_{i,j} p(i,j)^2 \quad (3.8)$$

Homojenlik, GSEM'deki elemanların dağılımının GSEM köşegenine yakınlığını ölçen $[0,1]$ aralığında bir değer döndürür. Aşağıdaki eşitlikte gösterildiği gibi hesaplanır.

$$Homojenlik = \sum_{i,j} \frac{p(i,j)}{1 + |i - j|} \quad (3.9)$$

Kontrast, bir imgedeki yerel gri seviye değişimlerini ölçer. Sabit imgelerde 0 değerini alır. Aşağıdaki gibi hesaplanır.

$$Kontrast = \sum_{i,j} |i - j|^2 \cdot p(i,j) \quad (3.10)$$

Kolerasyon, referans piksel ve komşu pikselin gri seviyede doğrusal bağımlılığını ölçer. $[-1, 1]$ aralığında değer alır. Kolerasyonun hesaplanabilmesi için öncelikle GSEM'in satır ve sütunlarına ilişkin ortalama ve standart sapma değerleri hesaplanmalıdır.

$$\mu_x = \sum_i \sum_j i p(i,j) \quad (3.11)$$

$$\mu_y = \sum_i \sum_j j p(i,j) \quad (3.12)$$

$$\sigma_x = \sum_{i,j} (i - \mu_x)^2 p(i,j) \quad (3.13)$$

$$\sigma_y = \sum_{i,j} (j - \mu_y)^2 p(i,j) \quad (3.14)$$

$$Kolerasyon = \sum_{i,j} \frac{(i - \mu_x)(j - \mu_y)p(i,j)}{\sigma_x \sigma_y} \quad (3.15)$$

Entropi imgedeki düzensizliği ölçer. Entropi ile enerji arasında ters bir ilişki vardır. Aşağıdaki gibi hesaplanır.

$$Entropi = \sum_{i,j} p(i,j) \cdot \ln p(i,j) \quad (3.16)$$

Çizelge 3.11 ve Çizelge 3.12’de KH2016 ve KH2017 veri setleri için gri seviyeli eş oluşturma matrisinden elde edilen doku özneliklerinin haploid ve diploid sınıflarına göre aldıkları *en küçük*, *en büyük*, *ortalama* ve *standart sapma* değerleri verilmiştir.

Çizelge 3.11 KH2016 veri seti doku özneliklerinin aldığı değerler

Öznitelikler	Haploid sınıfı				Diploid sınıfı			
	EK	EB	Ort	SD	EK	EB	Ort	SD
Enerji	0,9080	0,9523	0,9376	0,0079	0,8894	0,9721	0,9364	0,0092
Homojenlik	0,9956	0,9988	0,9979	0,0006	0,9939	0,9988	0,9972	0,0009
Kontrast	0,0322	0,0698	0,0479	0,0082	0,0033	0,0652	0,0323	0,0120
Kolerasyon	0,9679	0,9746	0,9724	0,0013	0,8657	0,9798	0,9676	0,0092
Entropi	0,1866	0,4149	0,2470	0,0380	0,1212	0,4411	0,2723	0,0410

EK: En küçük değer, EB: En büyük değer, Ort: Ortalama, SD: Standart sapma

Çizelge 3.12 KH2017 veri seti doku özneliklerinin aldığı değerler

Öznitelikler	Haploid sınıfı				Diploid sınıfı			
	EK	EB	Ort	SD	EK	EB	Ort	SD
Enerji	0,8823	0,9525	0,9311	0,0100	0,8318	0,9597	0,9232	0,0143
Homojenlik	0,9946	0,9992	0,9981	0,0006	0,9934	0,9995	0,9985	0,0006
Kontrast	0,0236	0,0797	0,0479	0,0084	0,0305	0,1376	0,0571	0,0111
Kolerasyon	0,9759	0,9859	0,9825	0,0014	0,9712	0,9859	0,9836	0,0009
Entropi	0,1862	0,5231	0,2807	0,0389	0,1492	0,4938	0,2740	0,0510

EK: En küçük değer, EB: En büyük değer, Ort: Ortalama, SD: Standart sapma

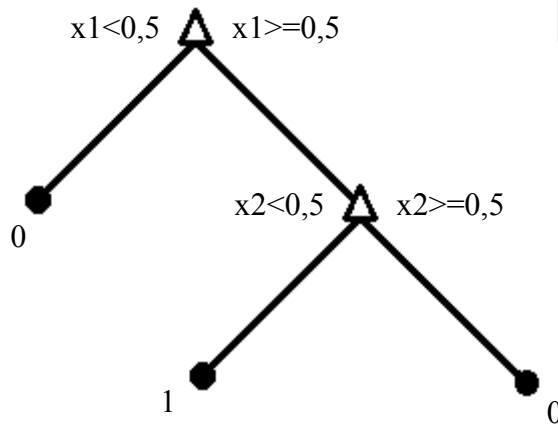
3.4. Sınıflandırma

Makine öğrenmesinde sınıflandırma, hangi sınıfa dahil olduğu bilinen örnekler üzerinden, hangi sınıfa dahil olduğu bilinmeyen ancak tahmine esas öznelik değerleri bilinen örnekler için sınıf tahmini yapmak için kullanılır [36]. Bu tez çalışması kapsamında, haploid ve diploid mısır tohumlarını R1nj markör sistemine göre sınıflandırmak için yaygın olarak kullanılan KA, k-EK, DVM ve YSA sınıflandırıcıları kullanılmıştır.

3.4.1. Karar Ağaçları

Karar ağacı yöntemi, bir ağaç yapısından esinlenerek sınıflandırma veya regresyon modelleri oluşturmak için kullanılır. Bir veri kümesini küçük ve daha da küçük alt kümelerle ayırırken, aynı zamanda ilişkili bir karar ağacı da aşamalı olarak geliştirilir. Elde edilen model ise karar düğümleri ve yaprak düğümleri olan bir ağaçtır. Bir karar düğümünün iki veya daha fazla dalı vardır. Yaprak düğümü bir sınıflandırmayı veya kararı temsil eder. Bir ağaçtaki en üstteki karar düğümü, kök düğüm adı verilen en iyi öngörücüye karşılık gelir. Karar ağaçları hem kategorik hem de sayısal verileri işlemekte yaygın olarak kullanılan bir yöntemdir.

Karar ağaçlarının oluşturulmasındaki en önemli adım ağaçtaki dallanmanın hangi öznelik değerlerine göre ağaç yapısının oluşturulacağıdır [37]. Literatürde bu problemin çözümü için geliştirilmiş gini indeksi, bilgi kazancı ve towing kuralı gibi çeşitli yaklaşımlar vardır. Bu tez çalışmasında, karar ağacı olarak gini indeksi kullanan CART algoritması [38] kullanılmıştır. Örnek bir karar ağacı modeli Şekil 3.14’de gösterilmiştir.



Şekil 3.14 Karar ağacı modeli

CART algoritması tarafından oluşturulan karar ağacı her zaman ikili bir karar ağacıdır (her düğümün yalnızca iki alt düğümleri olacaktır). CART algoritmasında karar ağacının oluşturulmasında kullanılan safsızlık (veya saflık) ölçüsü Gini İndeksi' dir (3.17). Diğer bir ifade ile CART algoritması için metrik olarak Gini İndeks parametresi kullanılmaktadır. Bu indeks eşitliğinde i ve j değerleri hedef değişkenin düzeylerini ifade etmektedir. Gini indeksi karar ağacı bölümlendirme işleminin niteliği hakkında bilgi verir. Bu indeks değeri bölünme sonrasında oluşan iki grupta sınıfların

ne kadar karmaşık olduğu hakkında bilgi verir. En iyi ayırmada bu değer 0 ile sonuçlanması iken, en kötü durumda ise 50/50 şeklinde sınıf bölümlenmesinin sonuçlanmasıdır.

$$Gini = \sum_{i \neq j} p(i)p(j) \quad (3.17)$$

3.4.2. k-en yakın komşuluk

K-en yakın komşu sınıflandırıcısı (k-EK), uygulaması kolay, öğrenme süreci güçlü, yaygın olarak kullanılan bir sınıflandırma yöntemidir. Eğitim setinde yer alan örnekler N özniteliğe sahipse, her bir örnek N -boyutlu uzayda bir nokta olarak temsil edilir. Sınıflandırılacak örnek, eğitim setinde yer alan örnekler ile arasındaki benzerliğe göre sınıflandırılmaktadır. Komşuluk olarak tanımlanan bu benzerlik, eğitim setindeki tüm örnekler ile sınıflandırılacak örnek arasındaki mesafenin hesaplanmasıyla belirlenir. Sınıflandırılacak örnek, en yakın k komşusunun sınıf etiketlerinin çoğunluğunu oluşturan sınıfa atanır. k değerinin seçimi sınıflandırıcının performansını doğrudan etkilemektedir. Doğru k değerinin seçimi için farklı yaklaşımlar vardır. Bunlardan biri farklı k değerleri için sınıflandırma performanslarını ölçüp en iyi performans gösteren k değerini seçmektir.

Komşulukların belirlenmesinde en yaygın olarak kullanılan mesafe ölçütü Öklid uzaklığıdır (3.18). Öklid uzaklığı dışında Minkowski uzaklığı, Manhattan uzaklığı, Chebyshev uzaklığı, Dilca uzaklığı da komşulukların belirlenmesinde kullanılan diğer mesafe ölçütleridir.

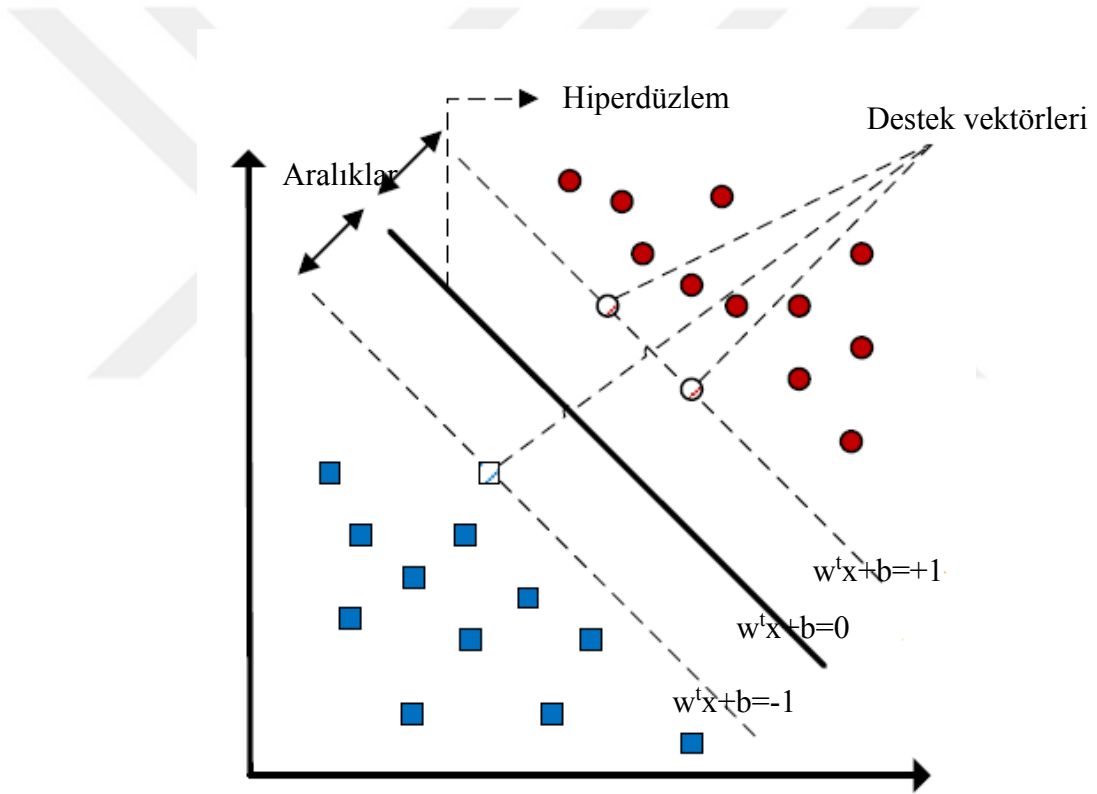
$$d_E = \sum_{i=1}^N \sqrt{(x_i - y_i)^2} \quad (3.18)$$

k-EK algoritması, örnek sayısının fazla olduğu eğitim setlerinde, üstün genelleme özelliği sayesinde başarılı sonuçlar verebilmektedir. Ancak sınıflandırılacak örneğin eğitim setindeki tüm örneklerle komşuluklarının hesaplanması, örnek sayısının fazla olduğu eğitim setlerinde hesaplama maliyetini arttırmaktadır [39]. Yöntemin performansını k en yakın komşu sayısı, eşik değer, benzerlik ölçümü ve öğrenme kümesindeki normal davranışların yeterli sayıda olması kriterleri etkilemektedir [40].

3.4.3. Destek vektör makinaları

Destek Vektör Makineleri (DVM) istatistiksel bir öğrenme algoritmasıdır. DVM ile sınıflandırma probleminin çözümünde temel prensip iki sınıfa ait örnekleri birbirinden optimum şekilde ayırabilen bir hiperdüzlemin belirlenmesidir [41]. Bu hiperdüzlem sınıflandırma hatalarını en aza indirmektedir.

DVM doğrusal olarak ayıramayan sınıflandırma problemlerinin çözümünde sınıflandırmaya esas karar kurallarını ifade eden hiperdüzlemlerin belirlenmesinde çekirdek fonksiyonları kullanılmaktadır [42]. Bu tez çalışmasında, DVM çekirdek fonksiyonu olarak radyal tabanlı fonksiyon kullanılmıştır. Şekil 3.15’de [43] DVM modeli gösterilmiştir.



Şekil 3.15 DVM modeli [43]

Bu çalışma kapsamında soft-margin DVM sınıflandırıcı kullanılmıştır. (Soft-margin) DVM sınıflandırıcısını hesaplamak, formun ifadesini en aza indirger. Bu amaçla (3.19) denklemini minimize etmeye çalışır. Bu denklemde λ margin boyu ile $\vec{x}_i$ parametresinin doğru tarafta olması arasında bir denge sağlar. Buradan yola çıkarak yeterli küçüklükte bir λ değeri ile kayıp fonksiyonunun ikinci terimi yok sayılabilir.

Bu yüzden, girdi verisi doğrusal olarak sınıflandırılabilir ise, hard-margin DVM gibi bir davranış sergileyecektir.

$$\left[\frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i(\vec{\omega} \cdot \vec{x}_i - b)) \right] + \lambda \|\vec{\omega}\|^2 \quad (3.19)$$

3.4.4. Yapay sinir ağları

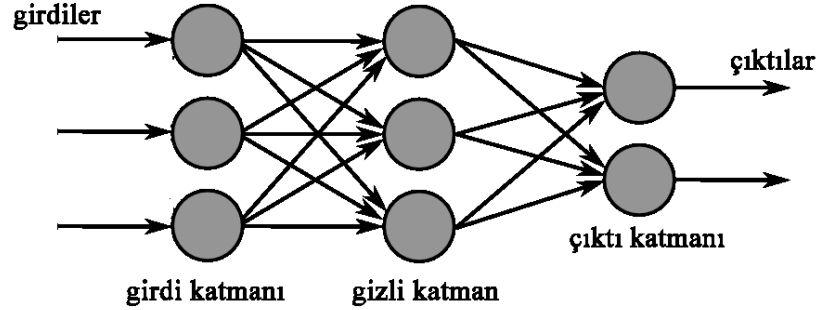
Yapay sinir ağları (YSA) ya da bağlantı sistemleri, hayvan beyinlerini oluşturan biyolojik sinir ağlarından esinlenerek geliştirilmiş bilgisayar davranışları modelleme sistemidir [44]. Bu tür sistemler, genel olarak göreve özgü programlama olmaksızın, örnekleri dikkate alarak "öğrenme" (yani performansı sürekli olarak geliştirerek) görevlerini öğrenirler.

Bir YSA, yapay nöronlar (biyolojik nöronların basitleştirilmiş versiyonu) olarak adlandırılan bağlı birimlerin veya düğümlerin bir setine dayanır. Yapay nöronlar arasındaki her bağlantı (sinaps yapısının basitleştirilmiş versiyonu) bir sinyali diğerine iletebilir. Sinyali alan yapay nöron onu işleyebilir ve ona bağlı yapay nöronları işaret edebilir.

YSA yaklaşımının asıl amacı, bir insan beyninin yaptığı gibi problemleri çözmektir. Bununla birlikte, zaman içerisinde literatürde belirli görevlerin eşleştirilmesine odaklanılmış ve bu durum biyolojiden esinli tasarımdan sapmalara yol açmıştır. YSA bilgisayarla görme, konuşma tanıma, makine çevirisi, sosyal ağ filtreleme, oyun tahtası, video oyunları ve tıbbi teşhis gibi çeşitli alanlarda kullanılmaktadır.

Bu çalışma kapsamında mısır görüntülerinden elde edilen özniteliklere göre mısırı haploid ya da diploid olarak sınıflandırmak amacıyla yapay sinir ağları kullanılmıştır. Yapay sinir ağlarında her bir katmanda ağırlıklandırılmış birbirine bağlı çok sayıda işlem biriminden (nöronlar) oluşan matematiksel sistemler yer alır. Bu işlemler arası geçişler ise transfer fonksiyonu olarak adlandırılan model ile sağlanmaktadır. Bu işlem birimlerinin temel görevi ise gelen girdiyi bir ağırlık seti ile ağırlıklandırmak çeşitli işlemlerle doğrusal olmayan bir şekilde dönüşümünü yapmak ve bu dönüşüm sonucundan da bir çıktı değeri üretmektir. Şekil 3.16'da çok katmanlı

örnek bir yapay sinir ağı verilmiştir. Bu sinir ağında birer adet giriş katmanı, gizli katman ve çıktı katmanı bulunmaktadır. Gizli katmanların sayısı modelde uygulanacak transfer fonksiyonlarının sayısına göre artırılabilir.



Şekil 3.16 Çok katmanlı örnek bir yapay sinir ağı

Matematiksel olarak bir nöron fonksiyonu $f(x)$ başka fonksiyonlara ayrışabilen $g_i(x)$ fonksiyonlarının birleşiminden (3.20) oluşur. Bu fonksiyonda K değeri aktivasyon (veya transfer) fonksiyonunu ifade etmektedir. Aktivasyon fonksiyonunun en önemli özelliği, giriş değerleri değiştikçe akıcı bir geçiş sağlamasıdır. Bu durumda girdi de yapılan küçük değişimler çıktıyı da küçük oranlarda etkiler. Aktivasyon fonksiyonu olarak hiperbolik tanjant, sigmoid fonksiyonu, softmax fonksiyonu ya da doğrultucu fonksiyonu gibi fonksiyonlar kullanılabilir.

$$f(x) = K \left(\sum_i \omega_i g_i(x) \right) \quad (3.20)$$

YSA' da en çok kullanılan transfer fonksiyonları tanjant-sigmoid (Tan-Sigmoid), logaritmik-sigmoid (Log-Sigmoid) ve Purelin fonksiyonlarıdır. Her bir fonksiyon girdilerin katmanlar arası geçişini modellemede kullanılır ve veri tipine göre farklı karakteristik özellikler sergilerler. Bu aktivasyon fonksiyonlarından Log-Sigmoid çok katmanlı sinir ağlarında geriye yayılım algoritması ile birlikte kullanılırlar. Log-Sigmoid 0-1 aralığında çıktı üretir. Tan-Sigmoid hiperbolik bir aktivasyon fonksiyonudur. Tan-Sigmoid -1, +1 aralığında bipolar bir çıktı üretir. Purelin aktivasyon fonksiyonu ise doğrusal olmayan (ancak nominal parametreler kullanıldığında doğrusala yakın) girdi/çıkı karakteristiği sergileyen modellerde kullanılır. Bu fonksiyon benzer şekilde -1, +1 aralığında doğrusal çıktılar üretir. Bu tez

çalışması kapsamında yaygın olarak kullanılan Tan-Sigmoid (3.21) fonksiyonu kullanılmıştır. Bu fonksiyonun temel avantajı hızlı çalışmasıdır.

$$tansig(n) = \frac{2}{1 + \exp(-2 * n)} - 1 \quad (3.21)$$

3.5. Performans Ölçümleri

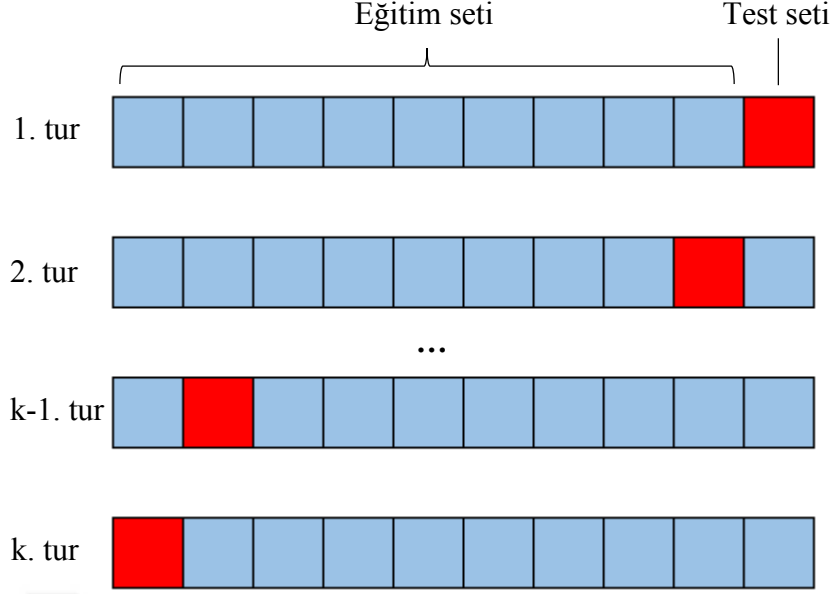
Bir sınıflandırma probleminde öznitelik çıkarımı, öznitelik seçimi ve bir sınıflandırıcının uygulanmasından sonra, önerilen modelin performansı ölçülmelidir. Bu ölçüm yapılırken veri seti, eğitim ve test seti olarak iki farklı kısma ayrılarak sınıflandırıcı başarımı değerlendirilir. Bu tez çalışması kapsamında performans ölçümünün güvenilirliğini artırmak için yaygın olarak bilinen k-katlı çapraz doğrulama yöntemi kullanılmıştır.

Sınıflandırma performansı test seti üzerinde çeşitli metrikler kullanılarak ölçülmektedir. Bu metrikler arasında hata matrisleri ve AİK eğrileri en yaygın olarak kullanılanlardır. Bu tez çalışmasında kullanılan sınıflandırıcıların performans ölçümlerinde hata matrisleri ve AİK eğrileri kullanılmıştır.

3.5.1. k-Katlı Çapraz Doğrulama

Sınıflandırma modellerinin eğitilmesi ve performans değerlendirmesi için veri setini eğitim ve test seti olarak parçalara ayırmak gerekir. Performans değerlendirmesi için veri setinin %70'ini eğitim, %30'unu test seti olarak kullanmak bir yaklaşımdır. Ancak burada veri parçalanırken verinin dağılımına bağlı olarak modelin eğitim ve testinde bazı sapmalar ve hatalar oluşabilir. Veriyi doğru temsil etmeyen bir model test setinde başarılı olabilir. Bu sebeple aslında başarılı olmayan bir model seçilen test setine bağlı olarak yüksek performans gösterebilir. Böyle bir yaklaşımda modelin farklı test setleri için performansları farklılıklar gösterecektir. Ayrıca örnek sayısının az olduğu veri setlerinde eğitim ve test setlerini ayırmak arzu edilmeyen bir durumdur.

k-katlı çapraz doğrulama, veriyi belirlenen bir k sayısına göre eşit parçalara böler, her bir parçanın hem eğitim hem de test için kullanılmasını sağlar, böylelikle dağılım ve parçalanmadan kaynaklanan sapma ve hataları asgariye indirir. Ayrıca veri setindeki tüm örnekler hem eğitim hem test için kullanılacağından örnek sayısı az olan veri setlerinin eğitim için az örnek kullanılmasını engeller.



Şekil 3.17 k-katlı çapraz doğrulama

Veri setinin eğitim ve test seti olarak nasıl seçileceği konusunda k -katlı çapraz doğrulama tekniği kabul görmüş bir yaklaşımdır. k -katlı çapraz doğrulamada, veri seti k parçaya ayrılır. Eğitim ve test işlemi k kez tekrarlanır. Her tekrarda, k parçalardan biri test seti olarak, diğer $k - 1$ parça ise eğitim seti olarak kullanılır. Daha sonra tüm k denemelerindeki ortalama hata hesaba katılır. Bu yöntem, verilerin eğitim ve test seti olarak ayrılması konusunun daha az önemsenmesini sağlamaktadır. Her veri örneği tam olarak bir kez test kümesine girer ve $k - 1$ defa eğitim kümesinde yer alır. Böylece veri setini oluşturan tüm örnekler bir kez test edilmiş olur.

3.5.2. Hata Matrisi ve Performans Metrikleri

Hata matrisi sınıflandırma probleminin çözümünde önerilen modelin başarımını ölçmek için kullanılan en temel ve yaygın kullanılan metriklerden biridir. İki veya daha fazla sınıflı sınıflandırma problemlerinde kullanılabilir.

İki sınıflı sınıflandırma problemlerinde, her ne kadar sınıf etiketi önemli olmasa da, arzu edilen sınıf pozitif (veya 0), diğer sınıf negatif (veya 1) olarak etiketlenir. Bu tez çalışmasında çözümü önerilen sınıflandırma probleminde haploidler pozitif, diploidler negatif sınıf olarak etiketlenmiştir.

Hata matrisi, iki boyutlu gerçek sınıf etiketi ve tahmin edilen sınıf etiketi ve her iki boyutta sınıf kümeleri olan bir tablodur. Gerçek sınıf etiketleri sütunları, tahmin

edilen sınıf etiketleri satırları oluşturur. Hata matrisi bir performans metriği değildir, ancak performans ölçümlerinin neredeyse tümü, hata matrisine dayanmaktadır. Şekil 3.18’de örnek hata matrisi gösterilmiştir.

		Gerçek Sınıf Değeri	
		Pozitif	Negatif
Tahmin Edilen Sınıf Değeri	Pozitif	DP	YP
	Negatif	YN	DN

Şekil 3.18 Hata matrisi

Hata matrisini oluşturan değerlerin tanımları aşağıda belirtilmiştir.

Doğru Pozitif (DP), Haploid sınıfına dahil olan ve haploid olarak tahmin edilmiş örnekleri ifade eder.

Doğru Negatif (DN), Diploid sınıfına dahil olan ve diploid olarak tahmin edilmiş örnekleri ifade eder.

Yanlış Pozitif (YP), Diploid sınıfına dahil olan ancak haploid olarak tahmin edilmiş örnekleri ifade eder.

Yanlış Negatif (YN), Haploid sınıfına dahil olan ancak diploid olarak tahmin edilmiş örnekleri ifade eder.

Hata matrisi sağladığı indisler sayesinde sınıflandırıcı karakteristiğini temsil eden çeşitli başarımların elde edilmesini sağlamaktadır. Bu kapsamda en sık kullanılan başarımların ölçütleri matematiksel formları ile birlikte aşağıda verilmiştir.

Doğruluk, sistemin tüm örnekler üzerindeki başarımını göstermektedir ve doğru tahminlerin tüm tahminlere oranı olarak ifade edilmektedir. Doğruluk metriği (3.22)'deki gibi tanımlanmaktadır.

$$\text{Doğruluk} = \frac{DP + DN}{DP + YP + DN + YN} \quad (3.22)$$

Duyarlılık, sistemin pozitif örnekler üzerindeki başarısını ölçmek üzere kullanılır ve (3.23)'deki gibi tanımlanır.

$$\text{Duyarlılık} = \frac{DP}{DP + YN} \quad (3.23)$$

Özgünlük, modelin negatif örnekler üzerindeki başarısını ölçmek üzere kullanılır ve (3.24)'deki gibi tanımlanır.

$$\text{Özgünlük} = \frac{DN}{DN + YP} \quad (3.24)$$

F-ölçütü (F1), duyarlılık ve özgünlük değerlerinin harmonik ortalaması olarak ifade edilir ve (3.25)'deki gibi tanımlanır.

$$F1 = \frac{2DP}{2DP + YP + YN} \quad (3.25)$$

3.5.3. AİK Eğrileri

Alıcı işletim karakteristiği (AİK) eğrisi, ikili sınıflandırma sistemlerinde ayırım eşik değerinin farklılık gösterdiği durumlarda, hassasiyetin kesinliğe olan oranıyla ifade edilmektedir. AİK daha basit anlamda doğru pozitiflerin, yanlış pozitiflere olan kesri olarak da ifade edilebilir. Her sınıflandırma işleminde yapıldığı gibi, metotlar kesinlik (yanlış pozitifleri eleme kabiliyeti) ve hassasiyet (doğru pozitifleri tespit etme kabiliyeti) arasındaki dengeyi kurmakla uğraşmaktadır. AİK eğrisi değişen sınıflandırma eşik değerlerine göre doğru pozitiflerin sayısının, yanlış pozitiflerin bir fonksiyonu olarak çizilmesiyle oluşmaktadır.

AİK eğrisinin altında kalan alan (AİK-AKA) sınıflandırıcının genelleştirme performansı hakkında bilgi verir. AİK-AKA değeri $[0, 1]$ aralığında değerler alır. AİK-AKA değeri 1 olduğunda, pozitifler mükemmel bir şekilde negatiflerden ayrılmış demektir. AİK-AKA değeri 0 olduğunda ise herhangi bir pozitif bulunmadığı anlamına gelir. İki sınıflandırıcı karşılaştırılırken AİK-AKA değeri büyük olan daha iyi bir genelleştirme performansı gösterdiği söylenebilir [43].

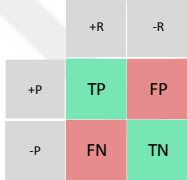
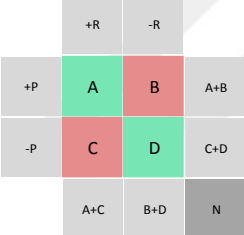


4. ARAŞTIRMA BULGULARI

4.1. Sınıflandırma Sonuçları

Verilerin eğitimi ve testi sırasında 10-katlı çapraz doğrulama yöntemi kullanılmıştır ve deneysel çalışma 50 defa tekrar edilmiştir. Hata matrisinde; TP ve TN sırasıyla doğru bir şekilde tespit edilmiş haploid ve diploid örnekleri temsil ederken; FP ve FN sırasıyla yanlış bir şekilde tespit edilmiş haploid ve diploid örnekleri temsil etmektedir. Sınıflandırıcı performanslarının değerlendirilmesinde kullanılan metrikler Çizelge 4.1’de gösterilmiştir.

Çizelge 4.1 Değerlendirme metrikleri [45, 46]

Hata Matrisi	Metrik	Formül	Açıklama
	Doğruluk	$\frac{A + D}{N}$	Doğruluk modelin genel performansını verir.
	Duyarlılık	$\frac{A}{A + C}$	Duyarlılık modelin hipoksik numuneler üzerindeki ayırt edici gücünü ortaya koymaktadır.
	Özgünlük	$\frac{D}{B + D}$	Özgünlük modelin ayırt edici gücünü normal örneklerde ölçer.
	F-ölçütü	$\left(\frac{2A}{2A + B + C} \right)$	F-ölçütü duyarlılık ve özgünlük arasındaki harmonik ortalamayı ifade eder.

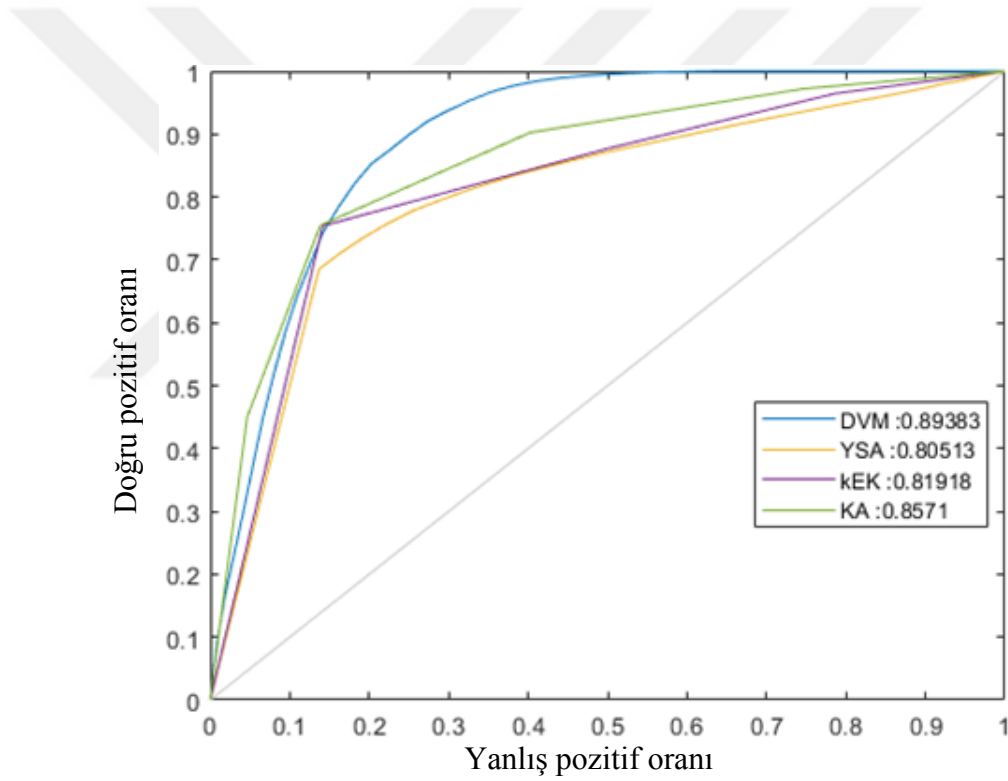
TP, FP, FN ve TN sayıları sırasıyla A, B, C ve D’ye karşılık gelir. + R (A + C), haploid örneklerin sayısını, -R (B + D) diploid örneklerin sayısını tanımlar. Benzer şekilde, + P (A + B), tahmin edilen haploid örneklerin sayısını, -P (C + D), tahmin edilen diploid örneklerin sayısını tanımlar. N örneklerin toplam sayısını temsil eder.

4.1.1. Renk histogramları

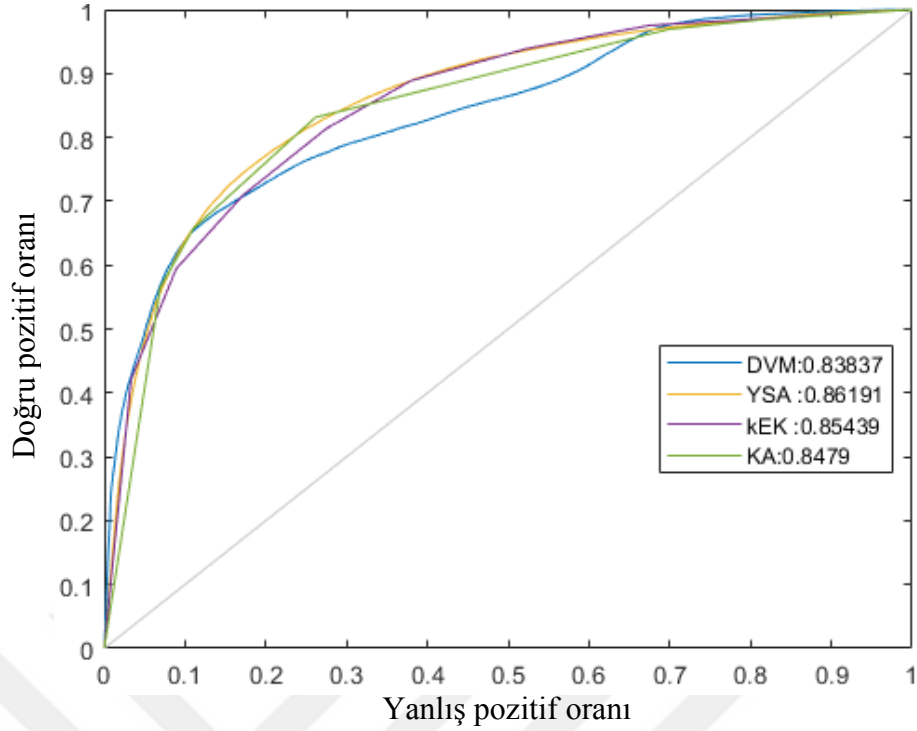
KH2016 ve KH2017 veri setleri üzerinde 8 renk ölçekli gri seviyeli renk histogramı öznitelik vektörlerinin sınıflandırılmasında KA bölünme kriteri gini index, en çok bölünme sayısı 7, en az yaprak sayısı 10 olarak; k-EK k değeri 7, uzaklık ölçüsü olarak öklid; DVM çekirdek fonksiyonu radyal tabanlı fonksiyon, sigma değeri 5, ceza parametresi 0,01 olarak; YSA eğitim algoritması trainlm, en çok iterasyon sayısı 500 ve ağ topolojisi {8, [12], 2} olarak ayarlanmıştır. Elde edilen sınıflandırma sonuçları Çizelge 4.2’de verilmiştir. KH2016 veri seti için sınıflandırıcı AİK eğrileri Şekil 4.1’de, KH2017 veri seti için sınıflandırıcı AİK eğrileri Şekil 4.2’de verilmiştir.

Çizelge 4.2 8 renk ölçekli gri seviyeli renk histogramı sınıflandırma sonuçları

	KH2016				KH2017			
	KA	k-EK	DVM	YSA	KA	k-EK	DVM	YSA
Doğruluk	84.93	79,68	75,53	80,25	79.98	77.40	78.43	80.22
Duyarlılık	62.14	52,36	92,21	56,79	68.57	73.04	68.97	71.89
Özgüllük	91.30	86,78	71,11	86,63	87.90	80.99	85.00	86.03
F-ölçütü	61.96	49,46	60,57	53,21	73.65	72.67	72.30	74.73
AİK-AKA	0.88	0,81	0,89	0,81	0.86	0.85	0.83	0.87



Şekil 4.1 KH2016 veri seti 8 renk ölçekli gri seviyeli renk histogramı öznelitektörünün sınıflandırıcı AİK eğrileri

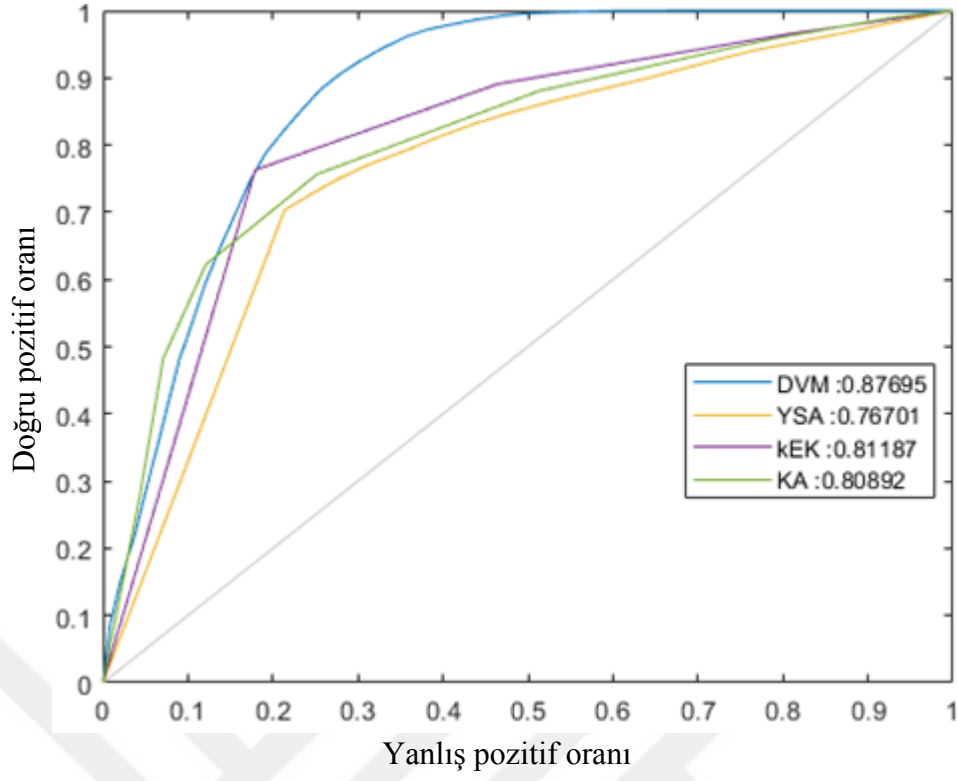


Şekil 4.2 KH2017 veri seti 8 renk ölçekli gri seviyeli renk histogramı öznitelik vektörünün sınıflandırıcı AİK eğrileri

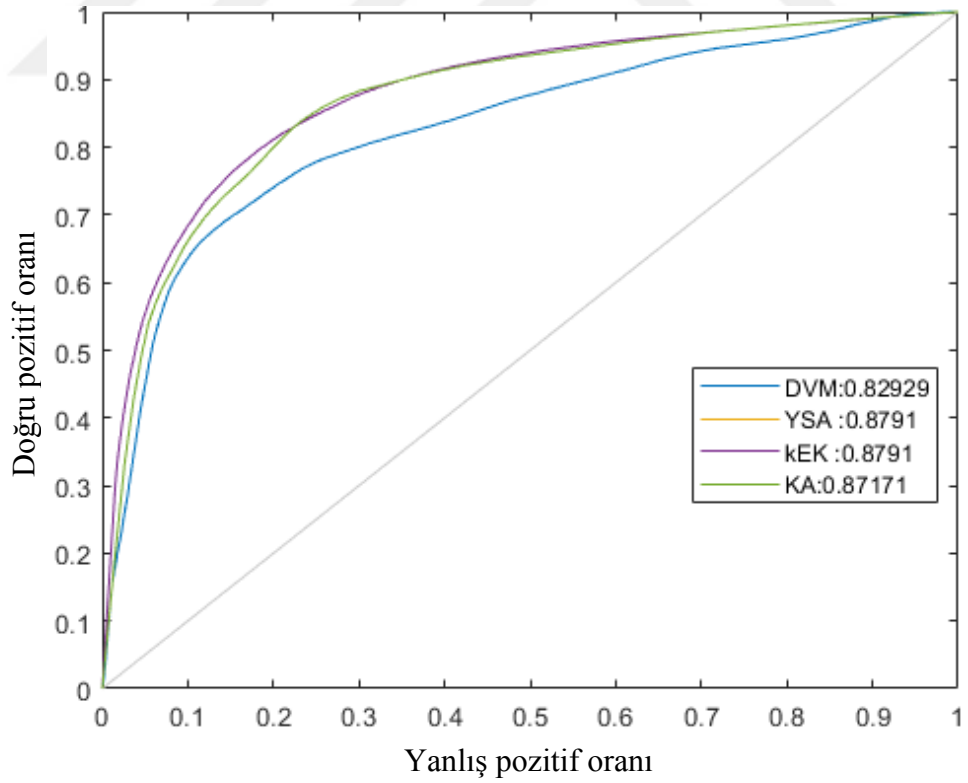
KH2016 ve KH2017 veri setleri üzerinde 16 renk ölçekli gri seviyeli renk histogramı öznitelik vektörlerinin sınıflandırılmasında KA bölünme kriteri gini index, en çok bölünme sayısı 7, en az yaprak sayısı 10 olarak; k-EK k değeri 7, uzaklık ölçüsü olarak öklid; DVM çekirdek fonksiyonu radyal tabanlı fonksiyon, sigma değeri 5, ceza parametresi 0,01 olarak; YSA eğitim algoritması trainlm, en çok iterasyon sayısı 500 ve ağ topolojisi {16, [12], 2} olarak ayarlanmıştır. Elde edilen sınıflandırma sonuçları Çizelge 4.3’de verilmiştir. KH2016 veri setinde sınıflandırıcı AİK eğrileri Şekil 4.3’de KH2017 veri setinde sınıflandırıcı AİK eğrileri Şekil 4.4’de verilmiştir.

Çizelge 4.3 16 renk ölçekli gri seviyeli renk histogramı sınıflandırma sonuçları

	KH2016				KH2017			
	KA	k-EK	DVM	YSA	KA	k-EK	DVM	YSA
Doğruluk	83.37	81,53	75,95	79,30	79.95	80.62	77.29	81.78
Duyarlılık	56.34	56,64	89,11	51,87	75.83	76.80	74.28	78.02
Özgüllük	90.64	87,74	72,50	86,70	82.82	83.27	79.41	84.40
F-ölçütü	57.02	53,39	60,10	50,08	75.56	76.14	72.77	77.77
AİK-AKA	0.81	0,81	0,87	0,77	0.88	0.88	0.82	0.88



Şekil 4.3 KH2016 veri seti 16 renk ölçekli gri seviyeli renk histogramı öznitelik vektörünün sınıflandırıcı AİK eğrileri

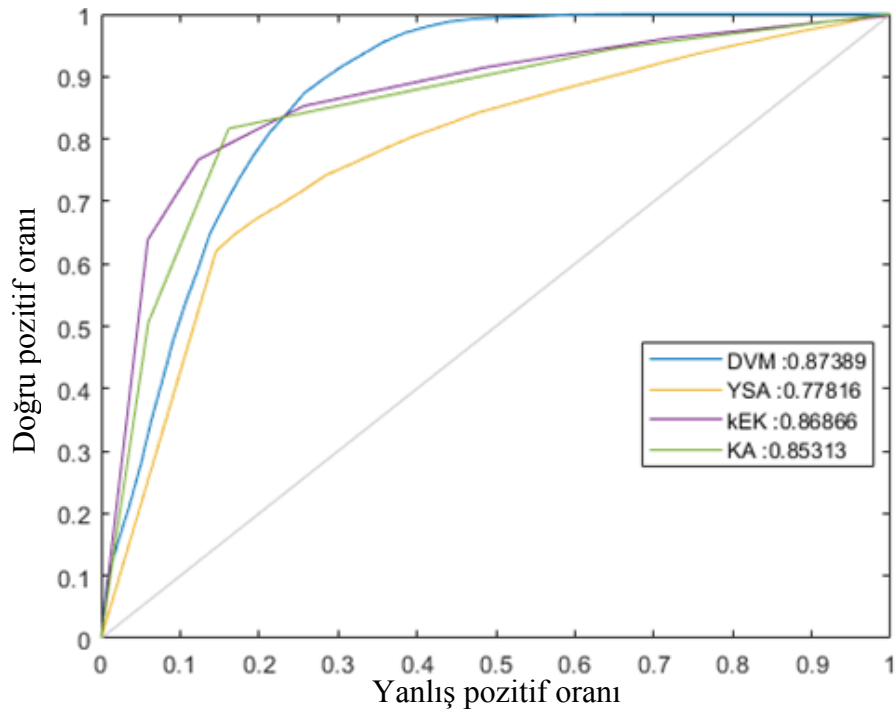


Şekil 4.4 KH2017 veri seti 16 renk ölçekli gri seviyeli renk histogramı öznitelik vektörünün sınıflandırıcı AİK eğrileri

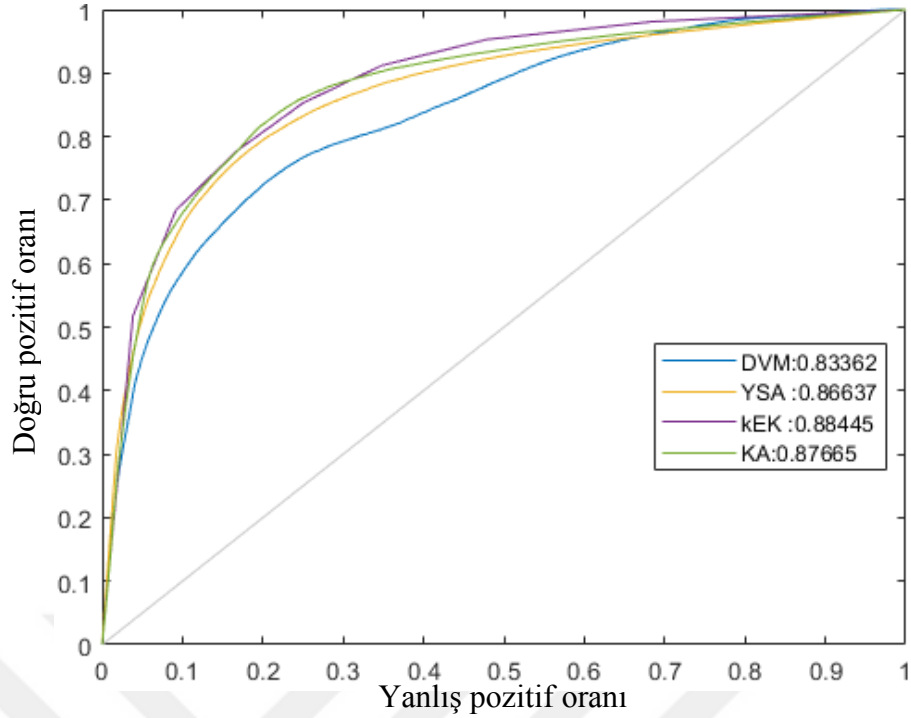
KH2016 ve KH2017 veri setleri üzerinde 32 renk ölçekli gri seviyeli renk histogramı öznitelik vektörlerinin sınıflandırılmasında KA bölünme kriteri gini index, en çok bölünme sayısı 7, en az yaprak sayısı 10 olarak; k-EK k değeri 7, uzaklık ölçüsü olarak öklid; DVM çekirdek fonksiyonu radyal tabanlı fonksiyon, sigma değeri 5, ceza parametresi 0,01 olarak; YSA eğitim algoritması trainlm, en çok iterasyon sayısı 500 ve ağ topolojisi {32, [12], 2} olarak ayarlanmıştır. Elde edilen sınıflandırma sonuçları Çizelge 4.4’de verilmiştir. KH2016 veri setinde sınıflandırıcı AİK eğrileri Şekil 4.5’de KH2017 veri setinde sınıflandırıcı AİK eğrileri Şekil 4.6’de verilmiştir.

Çizelge 4.4 32 renk ölçekli gri seviyeli renk histogramı sınıflandırma sonuçları

	KH2016				KH2017			
	KA	k-EK	DVM	YSA	KA	k-EK	DVM	YSA
Doğruluk	84.04	82,87	76,56	79,30	80.71	81.06	77.14	80.79
Duyarlılık	58.66	61,76	88,50	51,87	76.59	78.02	68.45	76.75
Özgüllük	90.88	87,62	73,44	86,70	83.59	83.06	83.20	83.62
F-ölçütü	58.82	54,30	60,46	50,08	76.45	76.39	70.99	76.54
AİK-AKA	0.85	0,87	0,87	0,77	0.88	0.88	0.83	0.87



Şekil 4.5 KH2016 veri seti 32 renk ölçekli gri seviyeli renk histogramı öznitelik vektörünün sınıflandırıcı AİK eğrileri



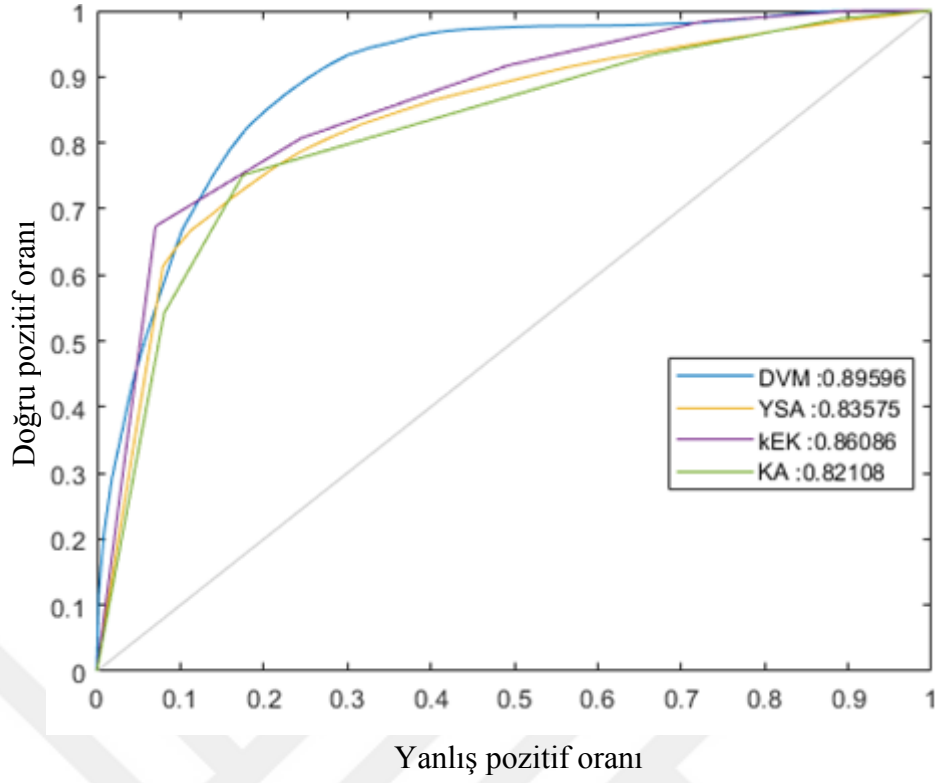
Şekil 4.6 KH2017 veri seti 32 renk ölçekli gri seviyeli renk histogramı öznitelik vektörünün sınıflandırıcı AİK eğrileri

4.1.2. Renk momentleri

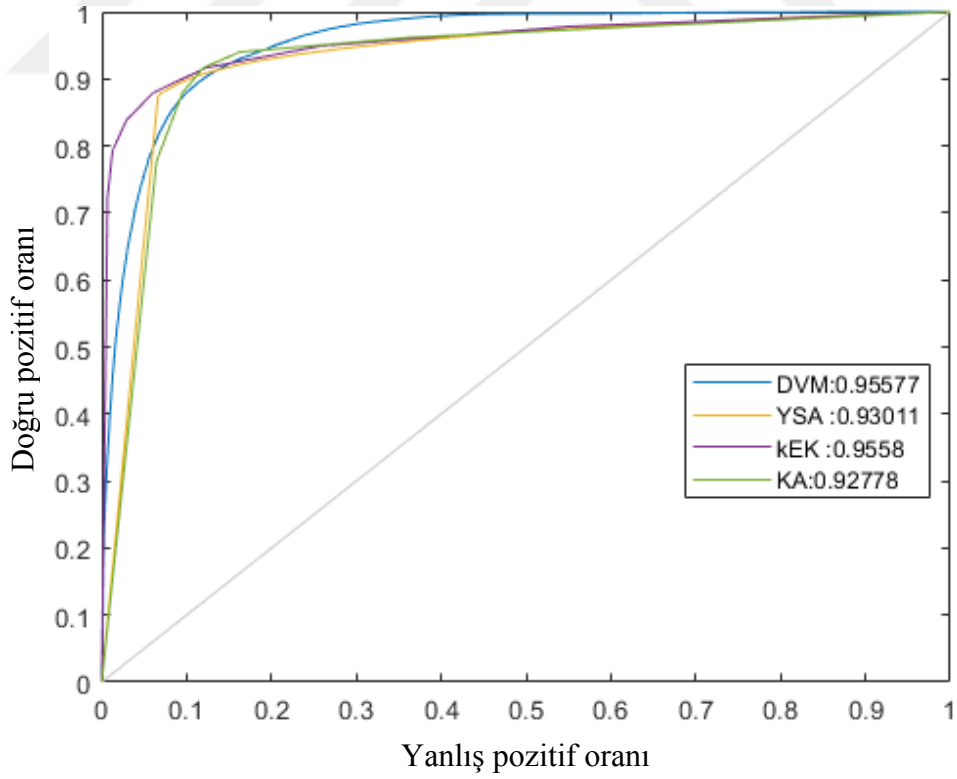
KH2016 ve KH2017 veri setleri üzerinde renk momentleri öznitelik vektörlerinin sınıflandırılmasında KA bölünme kriteri gini index, en çok bölünme sayısı 7, en az yaprak sayısı 10 olarak; k-EK k değeri 7, uzaklık ölçüsü olarak öklid; DVM çekirdek fonksiyonu radyal tabanlı fonksiyon, sigma değeri 5, ceza parametresi 0,01 olarak; YSA eğitim algoritması trainlm, en çok iterasyon sayısı 500 ve ağ topolojisi {12, [12], 2} olarak ayarlanmıştır. Elde edilen sınıflandırma sonuçları Çizelge 4.5’de verilmiştir. KH2016 veri setinde sınıflandırıcı AİK eğrileri Şekil 4.7’de, KH2017 veri setinde sınıflandırıcı AİK eğrileri Şekil 4.8’de verilmiştir.

Çizelge 4.5 Renk momentleri öznitelik vektörü sınıflandırma sonuçları

	KH2016				KH2017			
	KA	k-EK	DVM	YSA	KA	k-EK	DVM	YSA
Doğruluk	84.10	82,35	76,30	82,95	89.68	90.39	86.12	89.96
Duyarlılık	61.25	62,50	92,33	59,74	91.31	84.42	94.23	89.29
Özgüllük	90.73	87,41	71,95	88,37	88.55	95.40	80.51	90.43
F-ölçütü	60.53	54,28	61,23	57,28	87.85	88.88	84.73	87.90
AİK-AKA	0.82	0,86	0,89	0,82	0.92	0.95	0.95	0.92



Şekil 4.7 KH2016 veri seti renk momentleri öznitelik vektörünün sınıflandırıcı AİK eğrileri



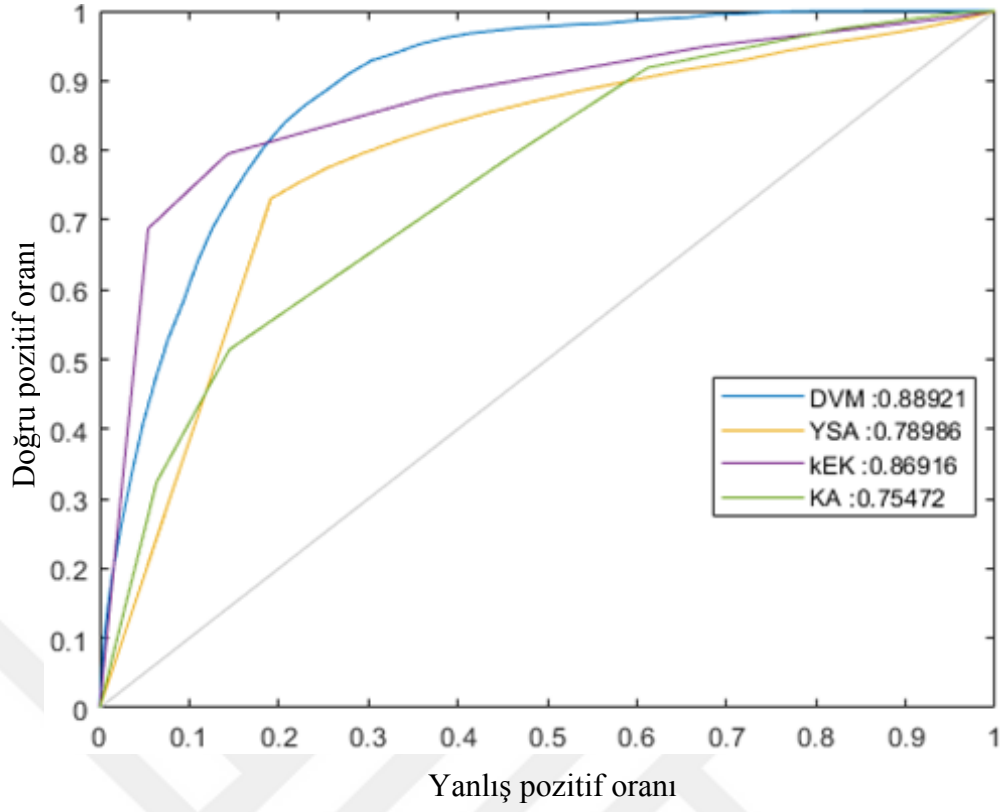
Şekil 4.8 KH2017 veri seti renk momentleri öznitelik vektörünün sınıflandırıcı AİK eğrileri

4.1.3. Doku öznitelikleri

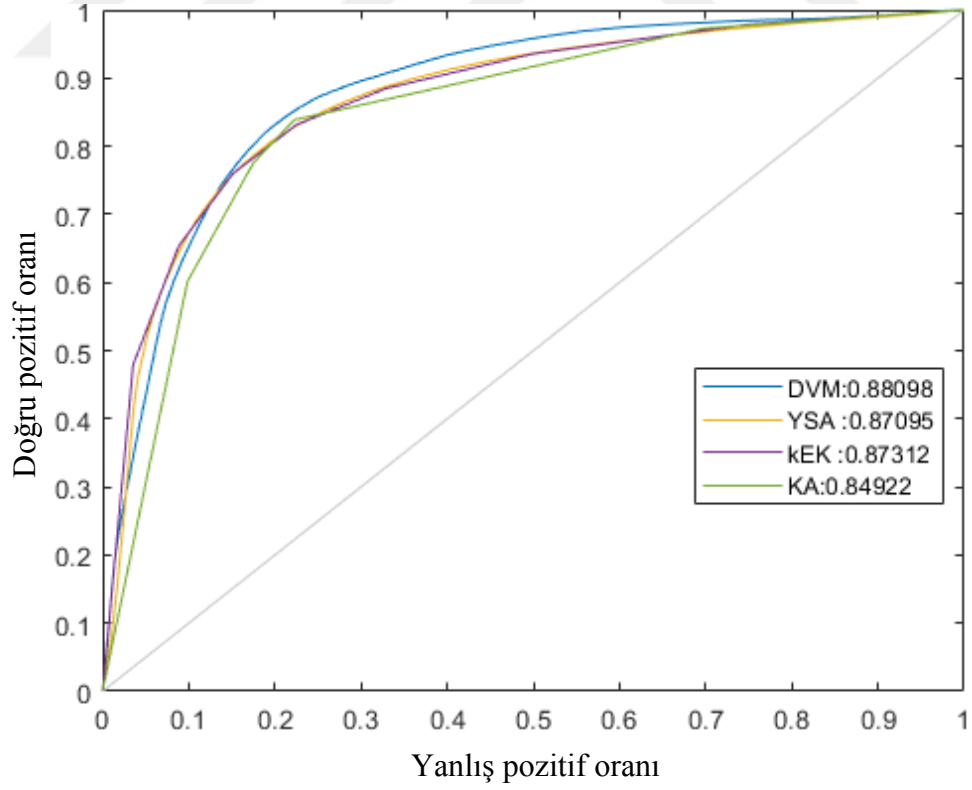
KH2016 ve KH2017 veri setleri üzerinde gri seviye eş oluşturma matrisinden elde edilen doku öznitelik vektörlerinin sınıflandırılmasında KA bölünme kriteri gini index, en çok bölünme sayısı 7, en az yaprak sayısı 10 olarak; k-EK k değeri 7, uzaklık ölçüsü olarak öklid; YSA eğitim algoritması trainlm, en çok iterasyon sayısı 500 ve ağ topolojisi {5, [12], 2} olarak ayarlanmıştır. DVM çekirdek fonksiyonu radyal tabanlı fonksiyon, ceza parametresi 0,01, KH2016 veri setinde sigma değeri 5, KH2017 veri setinde sigma değeri 2 olarak ayarlanmıştır. Elde edilen sınıflandırma sonuçları Çizelge 4.6’da verilmiştir. KH2016 veri setinde sınıflandırıcı AİK eğrileri Şekil 4.9’de, KH2017 veri setinde sınıflandırıcı AİK eğrileri Şekil 4.10’de verilmiştir.

Çizelge 4.6 Doku öznitelik vektörü sınıflandırma sonuçları

	KH2016				KH2017			
	KA	k-EK	DVM	YSA	KA	k-EK	DVM	YSA
Doğruluk	80.02	82,46	71,79	80,39	81.09	80.82	79.53	81.11
Duyarlılık	48.00	57,77	93,32	57,78	83.20	76.17	64.93	76.60
Özgüllük	88.88	89,81	65,27	86,51	79.64	84.17	89.70	84.26
F-ölçütü	47.72	58,48	58,20	54,01	78.23	76.74	72.14	76.81
AİK-AKA	0.75	0,87	0,89	0,79	0.85	0.87	0.88	0.87



Şekil 4.9 KH2016 veri seti doku öznelilik vektörünün sınıflandırıcı AİK eğrileri



Şekil 4.10 KH2017 veri seti doku öznelilik vektörünün sınıflandırıcı AİK eğrileri

5. TARTIŞMA VE SONUÇ

Bu tez çalışmasında, iki ayrı veri seti oluşturulmuştur. Veri setlerini oluşturan örneklerin sınıf etiketleri R1-nj renk markörüne göre iki konu uzmanı tarafından iki aşamada (görüntülerin elde edilmesinden önce ve sonra) manuel olarak belirlenmiştir. Hazırlanan veri setlerini oluşturan örneklerden 8, 16 ve 32 renk ölçekli gri seviyeli renk histogramları, renk momentleri ve GSEM doku öznetelikleri çıkarılmıştır. Böylece, toplamda 5 öznetelik vektörü elde edilmiştir. Elde edilen öznetelik vektörleri KA, k-EK, DVM ve YSA yöntemleri ile sınıflandırılmıştır. Sınıflandırıcı performansları 10-katlı çapraz doğrulama metodu 50 kez çalıştırılarak ölçülmüştür. Veri setlerini oluşturan örneklerin sınıfsal dağılımları dengesiz olduğu için sınıflandırıcı performanslarının karşılaştırılmasında F-ölçütü performans metriği dikkate alınmıştır.

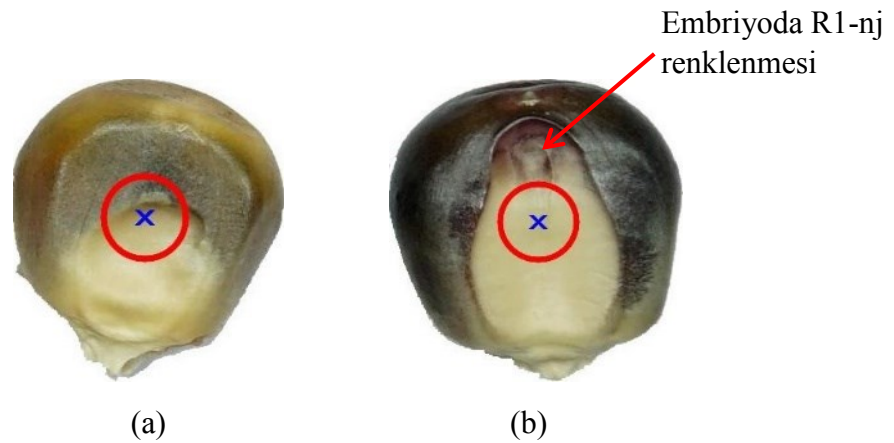
Yürütülen deneysel çalışmalar sonucunda KH2016 veri setinde 8 renk ölçekli gri seviyeli renk histogram öznetelik vektörünün sınıflandırılmasında en yüksek başarı KA'da duyarlılık (haploid mısır tohumu sınıflandırma başarımı) %62,14 ve özgüllük (diploid mısır tohumu sınıflandırma başarımı) %91,30 olarak ölçülmüştür. 16 renk ölçekli gri seviyeli renk histogram öznetelik vektörünün sınıflandırılmasında en yüksek başarı DVM'de duyarlılık %89,11 ve özgüllük %72,50 olarak ölçülmüştür. 32 renk ölçekli gri seviyeli renk histogram öznetelik vektörünün sınıflandırılmasında en yüksek başarı DVM'de duyarlılık %88,50 ve özgüllük %73,44 olarak ölçülmüştür. Renk momentleri öznetelik vektörünün sınıflandırılmasında en yüksek başarı DVM'de duyarlılık %92,33 ve özgüllük %71,95 olarak ölçülmüştür. GSEM doku öznetelik vektörünün sınıflandırılmasında en yüksek başarı k-EK'de duyarlılık %57,77 ve özgüllük %89,81 olarak ölçülmüştür.

KH2017 veri setinde 8 renk ölçekli gri seviyeli renk histogram öznetelik vektörünün sınıflandırılmasında en yüksek başarı YSA'da duyarlılık %71,89 ve özgüllük %86,03 olarak ölçülmüştür. 16 renk ölçekli gri seviyeli renk histogram öznetelik vektörünün sınıflandırılmasında en yüksek başarı YSA'da duyarlılık %78,02 ve özgüllük %84,40 olarak ölçülmüştür. 32 renk ölçekli gri seviyeli renk histogram öznetelik vektörünün sınıflandırılmasında en yüksek başarı YSA'da duyarlılık %76,75 ve özgüllük %83,62 olarak ölçülmüştür. Renk momentleri öznetelik vektörünün sınıflandırılmasında en yüksek başarı k-EK'de duyarlılık %84,42 ve özgüllük %95,40

olarak ölçülmüştür. GSEM doku öznelik vektörünün sınıflandırılmasında en yüksek başarı KA'da duyarlılık %83,20 ve özgüllük %79,64 olarak ölçülmüştür.

Fuente vd. [15] haploid tanıma için geliştirilecek bir modelin %100 başarı ile sınıflandırma yapmasının zorunlu olmadığını, %80'in üzerinde bir doğruluk oranının da emek ve para tasarrufu açısından oldukça faydalı olacağını belirtmişlerdir. Önerilen yöntemde en iyi performans, k-en yakın komşu sınıflandırma yöntemi ile KH2017 veri setinden elde edilen renk momentleri öznelikleri kullanılarak %90,39 doğruluk oranıyla elde edilmiştir. Bu doğruluk oranı önerilen yöntemin performansının tatmin edici düzeyde olduğunu göstermektedir.

Yanlış sınıflandırılan örnekler incelendiğinde embriyo bölgesinin bölütlenmesinde kullanılan kütle merkezli çemberin hatalı öznelikler çıkarılmasına sebep olduğu görülmüştür. Şekil 5.1'de [47] yanlış sınıflandırılmış haploid ve diploid mısır tohum örnekleri gösterilmiştir. Şekil 5.1 (a)'da haploid mısır tohumunda kütle merkezli çember embriyo ve endosperm bölgelerini içerdiğinden, endospermdeki R1-nj renklenmesi hatalı özneliklerin çıkarılmasına neden olmuş ve örnek yanlış sınıflandırılmıştır (yanlış negatif). Şekil 5.1 (b)'de diploid mısır tohumunda embriyoda R1-nj renklenmesi olmasına rağmen, renklenme kütle merkezli çemberin dışında kaldığından hatalı öznelikler çıkarılmış ve örnek yanlış sınıflandırılmıştır (yanlış pozitif).



Şekil 5.1 Yanlış sınıflandırılmış (a) haploid ve (b) diploid mısır tohum örnekleri

Haploid mısır tohumu tanıma konusunda araştırmacıların erişebileceği açık kaynak bir veri seti olmaması nedeniyle her araştırmacı kendi veri seti üzerinde çalışmalar yürütmektedir. Bu durum, çalışmaların karşılaştırılmasını engellemektedir.

Bu tez çalışması kapsamında oluşturulan veri setlerinde örneklerin sınıf etiketleri iki uzman tarafından iki aşamada manuel olarak yapılmıştır. Bu nedenle, veri setlerinde yanlış sınıf etiketine sahip örnekler olması muhtemeldir. Gelecekte kromozom sayımları yapıp ploidi seviyesine bakılarak sınıf etiketleri atanmış örneklerden oluşan bir veri seti inşa edilmesi planlanmaktadır. Bu sayede insan eliyle yapılan sınıf etiketi atamasından kaynaklı veri setindeki gürültü ortadan kaldırılmış olacaktır. Oluşturulacak veri seti bir sonraki adımda açık kaynak olarak web ortamında araştırmacıların kullanımına açılacaktır. Böylece konu üzerine çalışma yürütmek isteyen araştırmacılar standart bir veri seti üzerinde kendi algoritmalarını kıyaslama imkânı bulabilecektir.

Hatalı öznitelik çıkarılmasının önüne geçmek için embriyo bölgesini adaptif olarak bölütleyecek bir yöntemin geliştirilmesi de planlanmaktadır. Böylece elde edilecek özniteliklerin sınıflandırma performansında iyileştirme sağlayacağı düşünülmektedir. Ayrıca uçtan-uca (derin) öğrenme yöntemi ile öznitelik çıkarma işlemi olmaksızın yüksek doğruluk performansına sahip sistemlerin geliştirilmesi ile veri setlerinin sınıflandırılması planlanmaktadır. Son olarak, mısır ıslahçılarının haploid tanıma işlerini otomatik hale getirmek için gerçek zamanlı sınıflandırma yapacak bir makinenin tasarlanması amaçlanmaktadır.

KAYNAKLAR

- [1] C. Chidzanga, F. Muzawazi, J. Midzi, and T. Hove, *Production and use of haploids and doubled haploid in maize breeding : A review*, 4:4 (2017) 201–213.
- [2] Anonymous. (2018). <http://www.fao.org/faostat/en/?#data/QC> (on-line access on 2, Jul., 2018).
- [3] Anonymous. (2018). <https://biruni.tuik.gov.tr/medas/?kn=92&locale=tr> (on-line access on 2, Jul., 2018).
- [4] İ. Cerit, G. Cömertpay, R. Oyucu, B. Çakır, R. Hatipoğlu, and H. Özkan, *Melez Mısır Islahında In-Vivo Katlanmış Haploid Tekniğinde Kullanılan Farklı Inducer Genotiplerin Haploid İndirgeme Oranların Belirlenmesi*, **Tarla Bitk. Merk. Araştırma Enstitüsü Derg.**, 25:ÖZEL SAYI-1 (2016) 52–52.
- [5] H. H. Geiger, Doubled Haploids, *Handbook of Maize*. New York, NY: Springer New York, pp. 641–657, 2009.
- [6] P. Lashermes and M. Beckert, *Genetic control of maternal haploidy in maize (Zea mays L.) and selection of haploid inducing lines*, **Theor. Appl. Genet.**, 76:3 (1988) 405–10.
- [7] F. K. Röber, G. A. Gordillo, and H. H. Geiger, *IN VIVO HAPLOID INDUCTION IN MAIZE – PERFORMANCE OF NEW INDUCERS AND SIGNIFICANCE OF DOUBLED HAPLOID LINES IN HYBRID BREEDING*, .
- [8] B. M. Correctcitation and Prasanna, *Doubled Haploid Technology in Maize Breeding: Theory and Practice*, (2012).
- [9] A. Smelser, M. Blanco, T. Lübberstedt, A. Schechert, A. Vanous, and C. Gardner, *Weighing in on a method to discriminate maize haploid from hybrid seed*, **Plant Breed.**, 134:3 (2015) 283–285.
- [10] A. E. Melchinger, W. Schipprack, H. Friedrich Utz, and V. Mirdita, *In Vivo Haploid Induction in Maize: Identification of Haploid Seeds by Their Oil Content*, **Crop Sci.**, 54:4 (2014) 1497.
- [11] D. K. Nanda and S. S. Chase, *An Embryo Marker for Detecting Monoploids Of Maize (Zea Mays L.)*, **Crop Sci.**, 6:2 (1966) 213.
- [12] X.-Y. Wang, W.-X. Liao, D. An, and Y.-G. Wei, *Maize Haploid Identification via LSTM-CNN and Hyperspectral Imaging Technology*, (2018).
- [13] J. Lin, L. Yu, W. Li, and H. Qin, *Method for Identifying Maize Haploid Seeds by Applying Diffuse Transmission Near-Infrared Spectroscopy*, **Appl. Spectrosc.**, 72:4 (2018) 611–617.
- [14] P. Song, H. Zhang, C. Wang, B. Luo, and J. X. Zhang, *Design and Experiment of a Sorting System for Haploid Maize Kernel*, **Int. J. Pattern Recognit. Artif. Intell.**, 32:03 (2018) 1855002.
- [15] G. N. De La Fuente, J. M. Carstensen, M. A. Edberg, and T. Lübberstedt, *Discrimination of haploid and diploid maize kernels via multispectral imaging*, **Plant Breed.**, 136:1 (2017) 50–60.
- [16] B. W. Boote, D. J. Freppon, G. N. De La Fuente, T. Lübberstedt, B. J. Nikolau, and E. A. Smith, *Haploid differentiation in maize kernels based on fluorescence imaging*, **Plant Breed.**, 135:4 (2016) 439–445.

- [17] H. Qin *et al.*, *Identification of Haploid Maize Kernel Using NIR Spectroscopy in Reflectance and Transmittance Modes: A Comparative Study.*, **Guang Pu Xue Yu Guang Pu Fen Xi**, 36:1 (2016) 292–7.
- [18] J. Zhang, Junxiong and Wu, Zhanyuan and Song, Peng and Li, Wei and Chen, Shaojiang and Liu, *Embryo feature extraction and dynamic recognition method for maize haploid seeds*, **Trans. Chinese Soc. Agric. Eng.**, 29:4 (2013) 199–203.
- [19] E. G. de O. Couto, L. M. C. Davide, F. de O. Bustamante, R. G. Von Pinho, and T. N. Silva, *Identification of haploid maize by flow cytometry, morphological and molecular markers*, **Ciência e Agrotecnologia**, 37:1 (2013) 25–31.
- [20] L. W. Song Peng, Wu Kebin, Zhang Junxiong, Chen Shaojiang, Liu Jin, *Corn Haploid Grain Feature Extraction and Recognition*, **Trans. Chinese Soc. Agric. Mach.**, 3 (2012) 0–32.
- [21] C. Junxiong, Zhang and Kebin, Wu and Peng, Song and Wei, Li and Shaojiang, *Image segmentation of maize haploid seeds based on BP neural network*, **J. Jiangsu Univ. (Natural Sci. Ed.)**, 6 (2011) 000.
- [22] L. W. F. Junxiong, Song Peng¹ Wu Kebin¹ Zhang and Xianfa, *Sorting System of Maize Haploid Kernels Based on Computer Vision*, **Trans. Chinese Soc. Agric. Mach.**, (2010) S1.
- [23] W. Song, Peng and Zhang, Junxiong and Xun, Yi and Chen, Xiao and Li, *Development of automatic inspection system of corn seeds*, **Trans. Chinese Soc. Agric. Eng.**, 26:9 (2010) 124–127.
- [24] R. C. Gonzalez and R. E. (Richard E. Woods, *Digital image processing*, Prentice Hall, 2008 .
- [25] X. Chu, Y. Tao, W. Wang, Y. Yuan, and M. Xi, *Rapid Detection Method of Moldy Maize Kernels Based on Color Feature*, **Adv. Mech. Eng.**, 6 (2014) 625090.
- [26] Qiang Zhao, Jie Yang, Jingyi Yang, and Hongxing Liu, *Stone images retrieval based on color histogram*, 2009 International Conference on Image Analysis and Signal Processing, (2009) pp.157–161.
- [27] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack, *Efficient color histogram indexing for quadratic form distance functions*, **IEEE Trans. Pattern Anal. Mach. Intell.**, 17:7 (1995) 729–736.
- [28] D. Srivastava, R. Wadhvani, and M. Gyanchandani, *A Review: Color Feature Extraction Methods for Content Based Image Retrieval*, **IJCEM Int. J. Comput. Eng. Manag.**, 18:3 (2015) 2230–7893.
- [29] S. Sergyán, *Color Content-based Image Classification*, 5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics, (2007) pp.427–434.
- [30] K. Roy and J. Mukherjee, *Image Similarity Measure using Color Histogram, Color Coherence Vector, and Sobel Method*, **Int. J. Sci. Res.**, 2:1 (2013) 2319–7064.
- [31] Color Feature Extraction, *Perspectives on Content-Based Multimedia Systems*. Boston: Kluwer Academic Publishers, pp. 49–67, 2002.
- [32] F. Alamdar and M. Keyvanpour, *A New Color Feature Extraction Method Based on QuadHistogram*, **Procedia Environ. Sci.**, 10 (2011) 777–783.
- [33] R. M. Haralick, K. Shanmugam, and I. Dinstein, *Textural Features for Image Classification*, **IEEE Trans. Syst. Man. Cybern.**, SMC-3:6 (1973) 610–621.

- [34] R. M. Haralick, *Statistical and structural approaches to texture*, **Proc. IEEE**, 67:5 (1979) 786–804.
- [35] K. Hanbay and M. F. Talu, *Kumaş hatalarının online/offline tespit sistemleri ve yöntemleri*, **SAÜ Fen Bilim. Enstitüsü Derg.**, 18:1 (2014) 49.
- [36] S. B. Kotsiantis, *Supervised Machine Learning: A Review of Classification Techniques*, 31 (2007) 249–268.
- [37] Taşkın KAVZOĞLU and İsmail ÇÖLKESEN, *Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği Classification of Satellite Images Using Decision Trees: Kocaeli Case*, **Electron. J. Map Technol. Harit. Teknol. Elektron. Derg. Electron. J. Map Technol.**, 2:21 (2010) 36–4536.
- [38] L. Breiman, *Classification and Regression Trees*, Routledge, 2017 .
- [39] D. W. Aha, D. Kibler, and M. K. Albert, *Instance-Based Learning Algorithms*, **Mach. Learn.**, 6 (1991) 37–66.
- [40] Z. Cömert, A. F. Kocamaz, and V. Subha, *Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment*, **Comput. Biol. Med.**, (2018).
- [41] V. N. Vapnik, *The nature of statistical learning theory*, Springer, 2000 .
- [42] T. Kavzoglu and I. Colkesen, *Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi*, **Harit. Derg.**, :144 (2010) 73–82.
- [43] Z. Cömert, *Kardiyotokografi İşaretlerinin Analizi ve Makine Öğrenmesi Teknikleri ile Sınıflandırılması* , PhD Thesis, İnönü Üniversitesi, 2017.
- [44] E. Öztemel, *Yapay Sinir Ağları*, Papatya Yayıncılık, 2006 .
- [45] Z. Cömert and A. F. Kocamaz, *Comparison of Machine Learning Techniques for Fetal Heart Rate Classification*, **Acta Phys. Pol. A**, 132:3 (2017) 451–454.
- [46] Z. Cömert and A. F. Kocamaz, *Open-access software for analysis of fetal heart rate signals*, **Biomed. Signal Process. Control**, 45 (2018) 98–108.
- [47] Y. Altuntas, A. F. Kocamaz, R. Cengiz, and M. Esmeray, *Classification of haploid and diploid maize seeds by using image processing techniques and support vector machines*, 2018 26th Signal Processing and Communications Applications Conference (SIU), (2018) pp.1–4.

ÖZGEÇMİŞ

Ad Soyad: Yahya ALTUNTAŞ

Doğum Yeri ve Tarihi: Hekimhan - 1980

Adres: Kayısı Araştırma Enstitüsü Müdürlüğü, Aşağıbağlar mah. İsmetpaşa cad.
No:163 44090 Yeşilyurt / MALATYA

E-Posta: yahyaaltuntas@gmail.com

Lisans: Ege Üniversitesi, Bilgisayar Mühendisliği Bölümü (2004-2009)

Mesleki Deneyim ve Ödüller:

Yayın Listesi:

TEZDEN TÜRETİLEN YAYINLAR/SUNUMLAR

- Y. Altuntaş, A. F. Kocamaz, R. Cengiz and M. Esmeray, “Haploid ve diploid mısır tohumlarının görüntü işleme teknikleri ve destek vektör makinaları kullanılarak sınıflandırılması” *2018 26th Signal Processing and Communications Applications Conference (SIU)*, Izmir, 2018, pp. 1-4.