

**Kanserli Dokuların Mikroskop Görüntülerinde Kanser Kök Hücre Oranın Otomatik  
Olarak Belirlenmesi ve Klinikte Kullanılacak Yazılım Geliştirilmesi**

**Program Kodu: 1003**

**Proje No: 213E032**

Proje Yürütücüsü:

**Doç. Dr. Rengül ATALAY**

Araştırmacı:

Prof. Dr. Ahmet Enis Çetin

Doç. Dr. Çiğdem Gündüz Demir

Danışman:

Prof . Dr. Ayşegül Üner

Bursiyerler:

Can Fahrettin Koyuncu

Deniz Doğan

Diaa Badawi

Ece Akhan Güzelcan

Musa Tunç Arslan

Navid Mohammadvand

Oğuzhan Oğuz

Onur Yorulmaz

Rumeysa Fayetörbay

Simge Gökçe Örsçelik

Tülin Erşahin

## ÖNSÖZ

Kanser anormal hücrelerin kontrolsüz çoğalması ve yayılması olarak tanımlanan karmaşık bir hastalıktır. Bu nedenle kanser dokusu farklı özellikleri olan-ek popülasyon denilen-hücre gruplarını içerir. Bu gruplardan birisi olan kanser kök hücre popülasyonu kanserin kötü prognozundan sorumlu hücreler olarak tanımlanmaktadırlar. Hastalardaki hücre popülasyonunun birbirine göre oranı hastaların sağkalımında belirleyicidir.

Gerçekleştirilen bu proje kapsamında kanserin kemoterapötik tedavisinde izlenecek yol haritasında yönlendirici olması amacıyla kanserli dokularda tedaviye dirençliliğe yol açan CANSTEM kanser kök hücre oranını kanser doku kaynağına göre %76-90 oranında öngören yazılım geliştirilmiş ve patologların kullanımına sunulmuştur.

(<http://users.metu.edu.tr/rengul/canstem.html>). Bu proje 1003 program kodu ve 213E032 proje numarası ile TÜBİTAK tarafından desteklenmiştir.

## İÇİNDEKİLER

ÖNSÖZ .....	i
İÇİNDEKİLER .....	ii
TABLO VE ŞEKİL LİSTELERİ .....	iv
ÖZET .....	vii
ABSTRACT .....	viii
<b>1. GİRİŞ .....</b>	<b>1</b>
<b>2. LİTERATÜR ÖZETİ.....</b>	<b>2</b>
<b>3. GEREÇ ve YÖNTEM .....</b>	<b>4</b>
<b>3.1 Gereçler .....</b>	<b>4</b>
3.1.1 Dokuların görüntülerinin seçimi .....	6
<b>3.2 Metotlar .....</b>	<b>8</b>
3.2.1 İmmünohistokimya İmgelerinde Kanser Kök Hücrelerinin Bulunması .....	8
3.2.2 Betimleyici Karıştırma Metodu (BKM) ve CANSTEM yazılımı .....	11
3.2.3 CANSTEM Yazılımın Kullanıcı Arayüzü .....	14
3.2.4 H&E Boyanmış Dokularında Histopatolojik Kanser Kök Hücre İmge Sınıflandırılması .....	14
3.2.5 H&E ile Boyanmış Akciğer İmgelerinin Sınıflandırılması .....	18
<b>4. BULGULAR.....</b>	<b>19</b>
<b>4.1 H&amp;E ile Boyanmış Karaciğer İmgelerinin Sınıflandırılma Sonuçları .....</b>	<b>19</b>
<b>4.2 Akciğer İmge Seti Üzerinde Elde Edilen Deneysel Sonuçlar .....</b>	<b>28</b>
<b>4.3 Görüntü Analizi ve Dijital Patoloji Sempozyumu.....</b>	<b>40</b>
<b>5. TARTIŞMA VE SONUÇ .....</b>	<b>41</b>
<b>6. KAYNAKLAR LİSTESİ VE ATIFLAR .....</b>	<b>43</b>
<b>7. EK-1.....</b>	<b>45</b>

## TABLO VE ŞEKİL LİSTELERİ

### Tablo listesi

Tablo 1: Meme, Karaciğer ve akciğer kanseri kök hücre belirteçleri .....	3
Tablo 2: TMA-Setleri ve içerdikleri doku sayısı .....	4
Tablo 3: Seri kesit TMA doku camlarının içeriği .....	4
Tablo 4. Doku boyamaları için planlanmış olan ve test aşamasındaki antikorlar .....	5
Tablo 5: Kovaryans matrisine dayalı metotta çeşitli öznelik vektörleri kullanılarak elde edilen sonuçlar .....	20
Tablo 6: Ortak fark matrisine dayalı metodun sonuçları .....	20
Tablo 7 3-NN Kullanarak Elde Edilen Sınıflandırma Başarı Oranı .....	22
Tablo 8 5-NN kullanarak elde edilen sınıflandırma başarı oranı .....	22
Tablo 9 8-NN kullanarak elde edilen sınıflandırma başarı oranı .....	22
Tablo 10 10-NN kullanarak elde edilen sınıflandırma başarı oranı .....	22
Tablo 11 Ortak fark matrisi ve 3-NN kullanarak elde edilen sınıflandırma başarı oranı ..	23
Tablo 12 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi .....	25
Tablo 13 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	25
Tablo 14 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi .....	25
Tablo 15 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	26
Tablo 16 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi .....	26
Tablo 17 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	26
Tablo 18 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi .....	27
Tablo 19 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	27
Tablo 20 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi .....	27
Tablo 21 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	28
Tablo 22 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I-	

Derece II) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi .....	28
Tablo 23 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	28
Tablo 24 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi.....	29
Tablo 25 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	29
Tablo 26 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi.....	29
Tablo 27 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	30
Tablo 28 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi .....	30
Tablo 29 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	30
Tablo 30 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi.....	31
Tablo 31 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	31
Tablo 32 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi .....	32
Tablo 33 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	32
Tablo 34 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi .....	32
Tablo 35 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi .....	33
Tablo 36 Şekil 10'daki imgenin sonuçları .....	34
Tablo 37 Şekil 11'deki imgenin sonuçları .....	35
Tablo 38 CANSTEM Yazılımında kullanılan algoritma ile ImageJ renksel bölütleme eklentisinin karşılaştırılması .....	37
Tablo 39 CANSTEM'de kullanılan BKM ile EWB metotlarının karşılaştırılması .....	38

## Şekil listesi

Şekil 1. Seçilen örnek 3 meme TMA seri doku kesitlerinde EpCAM IHC boyaması gösterilmiştir. Camlar ışık mikroskobu altında 4X-10X-20X ve 40X büyütmelerde incelenmiş ve görüntüleri alınmıştır.....	6
Şekil 2. Aynı hastadan alınıp CD13 (solda) ve H&E (sağda) boyanmış seri kesim ardışık karaciğer doku katman imgeleri. ....	7
Şekil 3. Solda manuel etiketlenmiş immunohistokimya görüntü örneği. Sağda o bölgeye denk gelen H&E görüntüsü verilmektedir. ....	7
Şekil 4. Kanserli karaciğer dokusuna ait kanser kök hücreleri içeren bir CD13 imgesi....	8
Şekil 5. CD13 ile boyanmış bir kanserli imgenin (a), kahverengi bölgelerin işaretlendiği yer gerçekliği imgesi (b), ortak fark metodu (c), ve kovaryans metodu (d) ile elde edilen sonuçlar. En iyi sonuçlar ortak fark matrisi kullanıldığında elde edilmektedir. ....	11
Şekil 6. Eğitilen modellerin işlenen pencere üzerine verdikleri kararların LMS tabanlı algoritma ile birleştirilmesi .....	12
Şekil 7. CD13 ile boyanmış bir kanserli imgenin (a), yer gerçekliği imgesi (b), ilk sonuç (c), ve betimleyici karıştırma metoduyla eğitilmiş katsayılarla elde edilen sonuç (d). ....	13
Şekil 8. Aynı hastadan alınıp CD13 (solda) ve H&E (sağda) boyanmış ardışık karaciğer doku katman imgeleri. ....	15
Şekil 9: Sırasıyla Sınıf-1,Sınıf-2 ve Sınıf-3'e ait H&E ile boyanmış imgeler (20x büyütme). ....	18
Şekil 10. CD13 kanser kök hücre işaretleyicisiyle boyanmış karaciğer kanser dokusu .	35
Şekil 11. (a) CD13 ile boyanmış bir imge, (b) aynı imgenin kesin referans, (c) betimleyici karıştırma metodunun ilk sonucu, (d) kullanıcı geri dönüşü alındıktan sonraki sonuç. ...	36
Şekil 12. H&E ile boyanmış karaciğer kanseri imgesi. Koyu boyanmış çekirdekler büyük ihtimalle kanser kök hücreleridir. ....	38
Şekil 13. H&E ile boyanmış imgelerdeki koyu ve mavi çekirdek sayıları. Yatay ekseninde koyu çekirdek sayısı, dikey ekseninde ise sıradan bir şekilde maviye boyanmış çekirdeklerin sayısı vardır. Sınıf 1 CD13 ile boyanmış slaytlardaki KBO oranının %3'ten az olan imgeleri göstermektedir. Sınıf 2 ise %3'ten fazla olan imgeleri göstermektedir. Bu grafik oluşturulurken H&E slaytlarının 4 değişik bölgesinden 256X256'lık parçalar alınmıştır.....	39

## ÖZET

Kanser anormal hücrelerin kontrolsüz çoğalması ve yayılması olarak tanımlanan karmaşık bir hastalıktır. Bu nedenle kanser dokusu farklı özellikleri olan-ek popülasyon denilen-hücre gruplarını içerir. Bu hücrelerden olan kanser kök hücreleri normal kök hücreleri gibi kendini yenileme ve farklılaşma özelliklerini taşıırken normal kök hücrelerinin aksine homeostatik kontrolleri olmayan, yani farklılaşma özelliklerini dengeleyip, çevresel sinyallere göre programlama özellikleri olmayan hücrelerdir. Bu özellikleri dolayısıyla da son yıllarda kanser kök hücre popülasyonu kanserin kötü prognozundan sorumlu hücreler olarak tanımlanmaktadır. Bu bağlamda kanser tedavisinde kanser kök hücre oran bilgisi, hastaya verilecek tedavinin planlanması amacıyla kullanılabilir. Hekimler kanserli dokunun kanser kök hücre popülasyon oranına göre henüz kanser metastaz yapmamış bile olsa bireye özgü tedavi yaklaşımı izlenebilir.

Gerçekleştirilen bu proje kapsamında, kanserli parafin bloklara gömülü doku biyopsi örneklerinde, kanser kök hücre tanımlama öngörü aracını geliştirdik. Kovaryans matrisine dayalı sınıflandırıcılara ve Ana Bileşenler Analizi (PCA) algoritması sonucunda üç sınıf H&E boyanmış karaciğer kanseri dokularının sınıflandırma probleminde 76.0% imge sınıflandırma başarısı elde edilmiştir. Ana Bileşenler Analizi kanser hücresi belirleme probleminde %90'nın üstünde başarı ve kovaryans matrisine dayalı sınıflandırıcılar ile üç sınıf ayırıştırma probleminde %90'a yakın başarı sağlanmıştır.

Kanserli hastalarda bireye ve hedefe yönelik tedavi yol haritasının belirlenmesine destek olabilecek ve hastalığın tedavisinde ve prognozunda iyileştirme sağlayabilecek yazılımı hekimlerin kullanıma sunduk.

CANSTEM yazılıma <http://users.metu.edu.tr/rengul/canstem.html> adresinden ulaşılabilir.

**Anahtar Kelimeler:** Kanser kök hücre, makine ile öğrenme, immunohistokimya, tedavi, prognoz

## **ABSTRACT**

Cancer is a complex disease characterized by uncontrolled cell proliferation and metastasis. For this reason, cancer tissue includes cell groups called side-population, which have different properties. Similar to the normal stem cells, cancer stem cells possess self-renewing and differentiation capacities but lack homeostatic controls. Cancer stem cells don't respond to the external which control growth regulation and differentiation. Because of these features, cancer stem cells in recent years have been described as cell populations responsible for the poor prognosis of cancer. In this context, cancer stem cell ratio information can be used to plan the treatment. Physicians can monitor the individualized treatment approach even if the cancer has not yet metastasized at early stage.

Within the scope of this project, we developed a cancer stem cell prediction tool in cancerous tissue biopsy specimens embedded in paraffin blocks. Classification based on the covariance matrix and PCA algorithm resulted in 76.0% image classification success for the classification H&E stained liver cancer tissues. The Principal Component Analysis achieved over 90% success and the covariance matrix in the classification problem of three classes had a success nearly 90% in cancer detection.

The newly developed CANSTEM software allows the identification cancer stem cell ratio in cancer tissues for individualized and targeted treatment and prognosis of the disease. CANSTEM software is available at <http://users.metu.edu.tr/rengul/canstem.html>.

**Key words: Cancer stem cell, machine learning, immunohistochemistry, treatment, prognosis**



# 1. GİRİŞ

Kanser dünya sađlık örgütü verilerine göre, kardiyovasküler hastalıkların da önüne geçerek en ölümcül hastalık olarak sıralanmıştır. Kanser olgularında modern tedavi yaklaşımlarına bile dirençlilik, metastaz ve nüks (tekrar) bu hastalığın kötü prognozunun en önemli nedenlerdir. Kanser dokusu oldukça heterojen bir yapıya sahip olup deđişik hücre alt gruplarını içerir. Kanserli hücrelerin genomlarındaki hasarın hemen hemen her hücrede birbirinden farklı bir yapıya sahip olması bu hücrelerin kemoterapiye verdiği yanıtta da farklılıklar yaratmaktadır. Son yıllarda yapılan çalışmalar, bu alt gruplar arasında yer alan kanser kök hücrelerinin tedaviye dirençlilik metastaz ve nüks açısından önemli rol oynadıklarını göstermektedir. Bu nedenle patolojik dokulardaki kanser kök hücre oranının saptanması hastalığın tedavisinde izlenecek yol ve prognozu açısından çok önemlidir. Böylece tedavi verecek olan hekim bu bilgiyi kullanarak kanserli hastada kişiye özgü bir protokol izleyebilir.

Hedeflerine proje önerisinde öngörüldüğü üzere kanserli dokulardan elde edilen rutin patolojik Hematoksilen-Eozin Boyama (HxE) mikroskop görüntülerinde kanser kök hücre oranının görüntü analizi yöntemleri ile otomatik olarak belirlenmesi ve klinikte kullanılacak bir yazılım geliştirilmiştir. Dokuya kanser tanısını veren klinik patoloğlar rutin HxE doku boyamasında kanser kök hücre oranını belirleyemezler. Bunun için özel kanser kök hücre belirteçleri olan antikor belirteçleri yardımı ile immunohistokimyasal yöntemler uygulanabilir ancak klinik rutinde kanser kök hücre belirteçlerine dayalı bir uygulama bulunmamaktadır. Bu nedenle dokudaki kanser kök hücreleri belirleyen bu uygulama hem zaman hem de ekonomik açıdan rutin olarak kullanılmamaktadır. Proje kapsamında geliştirilen CANSTEM (<http://users.metu.edu.tr/rengul/canstem.html>) öngörü aracı, rutin patolojik preparatlarda kanser kök hücre oranını klinik patoloğ ve onkoloğ hekimlerin bilgisine 42. Ulusal Hematoloji Kongresinde (19-22 Ekim 2016) sunulmuştur. CANSTEM kanserli hastaların tedavisinde kişiye özgü yol haritası oluşturmalarını sağlayacak şekilde tasarlanmıştır. Böylece CANSTEM tedavinin gecikmeden en etkin bir şekilde gerçekleştirilmesi gereken kanser hastalığında hem hesaplı hem de hızlı bir şekilde kullanılabilir. Onkoloğlar ve patoloğlar uzun vadede öngörü aracını kullanarak kanserli hastalarda kişiye özgü kanser kök hücrelerini hedef alan tedavi yol haritası ile bu ölümcül hastalığın tedavisinde ve prognozunda iyileştirme sağlayabileceklerdir.

Alt konu başlıkları sola dayanmalı, sadece iki rakam arasına nokta konmalıdır. Sözcükler büyük harfle başlamalı ve koyu tonda yazılmalıdır. İkiiden çok alt başlık kullanmaktan kaçınılmalıdır.

## 2. LİTERATÜR ÖZETİ

Çoğu kanser çeşidinde bulunduğu rapor edilen kanser kök hücre olarak tanımlanan hücrelerin varlığının, kanserin gelişiminde ve metastazında da rol aldığı gösterilmiştir (Correnti ve Raggi, 2016). Genel olarak kanser kök hücreleri, normal kök hücreleriyle ortak yönleri olmasının yanı sıra bazı özellikleri ile farklılık gösteren hücrelerdir. Kanser kök hücreleri normal kök hücreler gibi kendini yenileme ve farklılaşma özelliklerini buldurmalarına rağmen hücre döngülerini homeostatik bir şekilde kontrol edemez ve hücre çoğalmasını kontrol eden negatif sinyallere olumlu yanıt veremezler. Kanser kök hücrelerinin başlıca üç temel özellikleri aşağıdaki gibi sıralanabilir: i) Tümör hücrelerinin sadece azınlıkta olan bir kısmı tümörleşme potansiyeline sahiptir ; ii) Tümör oluşturu hücreler, hücre yüzeylerinde diğer hücrelerden farklı olarak bulunan moleküllerin varlığı ile karakterize edilebilirler; ve iii) Tümör oluşturabilen popülasyondan üretilen tümörlerin karışık kanser hücresi popülasyonlarının oluşmasına sebep verdiği dolayısıyla primer tümörün heterojen fenotipini tekrardan yaratmaktadır (Dalerba vd., 2007). Tümör hücrelerinin heterojenik yapısı, kanser ilaçlarına olan direncin gelişmesindeki en önemli sebeptir<sup>3</sup>.

Tümör hücrelerinin hem dirençliliğinden hem de metastazından sorumlu olarak görülen kanser kök hücrelerinin, epitelial-mezenkimal geçişi (EMT) ve normalde embriyolojik gelişim evresinde farklı dokuların oluşmasında görev alan mekanizmaları, kanser hücrelerinde de çeşitli hücre yolakları (örneğin; TGF- $\beta$ ) üzerinden indükleyerek aktif hale getirerek tümörün yayılmasına neden olduğu savunulmaktadır. Aktif hale gelen hücre yolaklarının sebep olduğu olaylar, epitel hücrelerde var olan E-cadherin proteininin ifadesinin bastırılmasını sağlayan proteinlerin (Snail, Slug ve Zeb1 gibi) ifadesinin artmasına yol açar (Peinado vd., 2007). Ayrıca, Kanser hücre hatlarında EMT'in karakterize edilmesine dayanan çalışmalarda, E-cadherin proteinini ifade eden ve vimentin proteinini ifade etmeyen hücre hatları epitelial; E-cadherin proteinini ifade etmeyip, vimentin proteini yüksek seviyede ifade edenler ise mezankimal olarak sınıflandırılmaktadır. Bu bağlamda hücrelerin EGF reseptörü inhibitörlerine olan duyarlılığı EMT ile ilişkilendirilebilmiştir. Epitelial özellikte olan hücrelerin mezankimal olanlara göre EGF reseptörü inhibitörlerine daha duyarlı olduğu saptanmıştır. Kanser

kök hücrelerinin uygulanan konvansiyonel tedavi yöntemlerine karşı direnç kazanmasının altında bu hücrelerin kolaylıkla mikro çevresinde değişiklikler yaratabilip kendini koruma altına alabilme özelliği taşımaları yatar. Bu nedenle, kanser kök hücrelerini yok etmek için klasik tedavi yöntemlerinin yansıra kanser kök hücrelerinin sağkalım mekanizmalarının hedeflenmesinin uygun olacağı önerilmektedir (Oishi ve Wang, 2011). Bu mekanizmalardan ilki kanser kök hücresinde aktif olan hücre yolaklarını bloke etmek, yani hücrelerin kendini yenilerken, tümörün gelişimini tetiklerken ve farklılaşırken indüklediği yolları hedeflemektir (Wnt/ $\beta$ -catenin, Hedgehog, PTEN gibi). İkincisi kanser kök hücrelerinin yüzeyinde bulunan belirteçlerin (CD133, CD13, EpCAM, CD44) hedeflenmesi ve bu şekilde alt sinyal yollarının indüklenmesinin engellenmesidir. Üçüncüsü tümör mikro-çevresinin bozuluma uğratılması yöntemidir. Bu yöntemde damarlaşmayı, metastazı ve hipoksiyi tetikleyen faktörlerin inhibe edilmesi hedeflenir. Son olarak kanser kök hücrelerinin kendini koruma mekanizmalarını (bağışıklık sisteminden kurtulma, ilaca karşı direnç veya radyoterapi direnci) bozmak hedeflenebilir. Ancak her yaklaşım için dokudaki kanser kök hücre oranının belirlenmesi gerekmektedir.

Yukarıda belirtildiği gibi kanser dokusunun moleküler özellikleri ve tedaviye yanıtı çok farklılık göstermektedir. Sağlık Bakanlığı Kanser Daire Başkanlığı ve GLOBOCAN verilerine göre ülkemizde en sık görülen akciğer, meme ve karaciğer kanserleri görülmektedir<sup>6</sup>. Kanser kök hücrelerinin tanımlanmasında değişik biyobelirteç antikolar literatürde tespit edilmiştir (Tablo. 1) (Iqbal vd., 2013; Liu vd., 2011; O'Flaherty vd., 2012)

**Tablo 1: Meme, Karaciğer ve akciğer kanseri kök hücre belirteçleri**

Meme	Karaciğer	Akciğer	
		Küçük hücreli	Küçük hücreli olmayan
CD44	CD133	CD13	CD133
CD44+/CD	CD44	ALD	
ALDH1	EpCAM	CD16	
CD133	CD13	CD44	
	OV		

Literatürde tespit edilen bu antikor biyobelirteçlerden temin edilebilenleri immünohistokimya yöntemleri için proje kapsamında kanser kök hücre tespitinde kullanılmıştır.

### 3. GEREÇ ve YÖNTEM

#### 3.1 Gereçler

Proje kapsamında akciğer karaciğer ve meme kanserler için seri kesim doku örnekleri BioMAX (<http://www.biomax.us/tissue-arrays>), USA firması ile iletişime geçilerek temin edilmiştir (Tablo 2). 190 Akciğer, 131 Karaciğer ve 247 Meme TMA preparatının doku seri doku kesit içeriği (Tablo 3, de belirtilmiştir. Proje kapsamında kullanılan dokular hem kanser hem de normal veya diğer hastalık dokularını içermektedir. Bu dokular proje yazılım geliştirme sürecinde negatif kontrol olarak kullanılmıştır.

**Tablo 2: TMA-Setleri ve içerdikleri doku sayısı**

Akciğer		Karaciğer		Meme	
TMA kodu	Doku sayısı	TMA kodu	Doku sayısı	TMA kodu	Doku sayısı
LC817	81	LVC961	96	BR8017	81
LC952	48	LV962	97	BR963a	97
LC814	81	LV8011	81	BR953	48
Toplam doku kesiti	<b>210</b>		<b>274</b>		<b>226</b>

**Tablo 3: Seri kesit TMA doku camlarının içeriği**

Temin edilen miktar	TMA cam kodu	Camda yer alan tümör örneği sayısı	Camda yer alan normal doku örneği sayısı	Organ ile ilgili diğer patolojiler
6	LC814 - Akciğer	39	0	41
6	LC817- Akciğer	40	0	41
6	LC952 (ikili)- Akciğer	30	8	9
6	LV962-Karaciğer	66	6	27
6	LVC961- Karaciğer	58	6	24
6	LV8011- Karaciğer	49	5	20
6	BR8017-meme	74	8	1
6	BR963a-meme	76	2	13
5*	BR953 (ikili)-meme	32	8	10

\* BioMAX firmasında sadece 5 cam seri kesit doku için örnek bulunmaktaydı.

Daha sonra seçilen TMA'lar HxE veya IHC boyanmadan önce Tablo 3'te belirtilen KKH biyobelirteci olan proteinlere karşı reaktif antikorlar temin edilmiştir. Proje danışmanımız Prof.

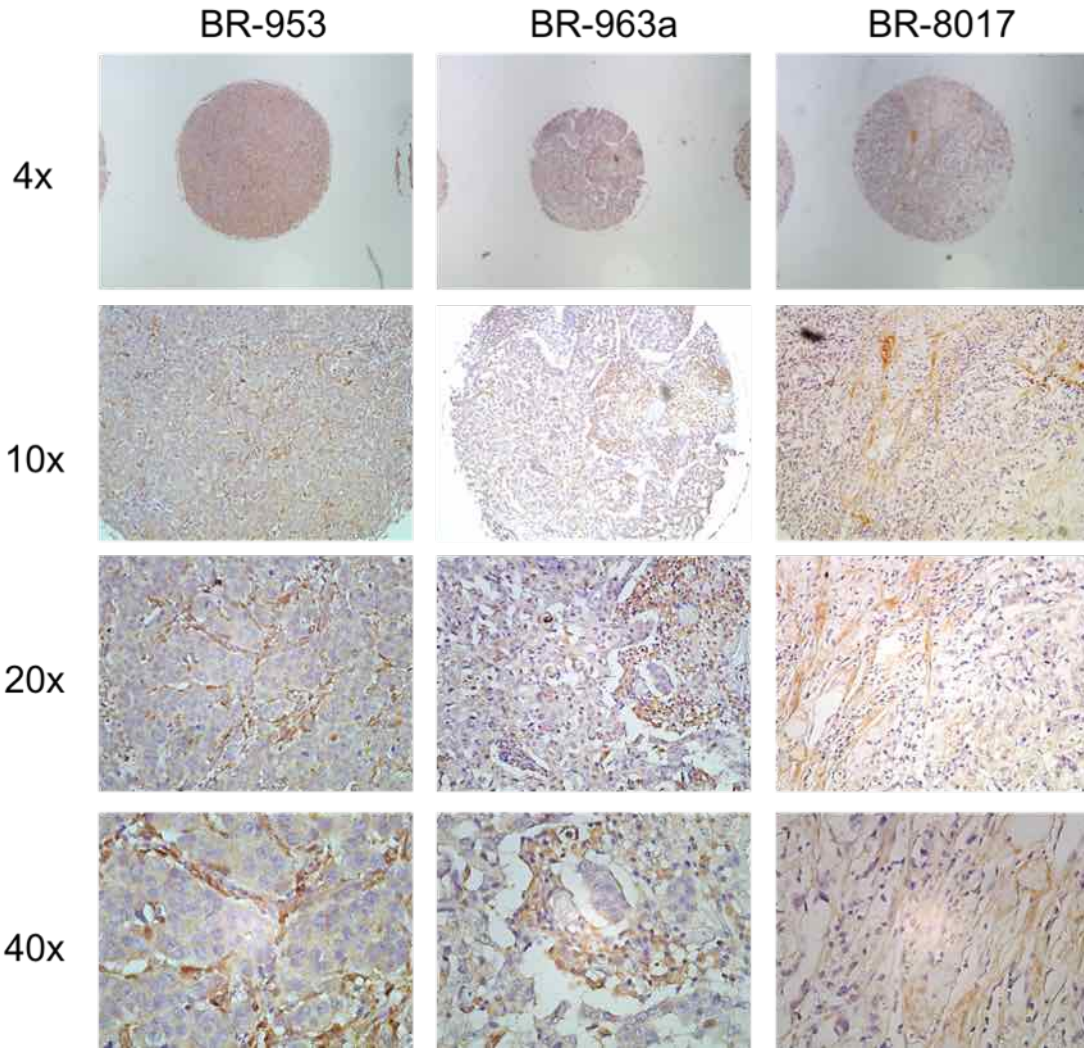
Dr. Ayşegül Üner rehberliğinde bu antikorların IHC reaktiviteleri Hacettepe Üniversitesi Patoloji bölümünden temin edilen anonim kanser dokularında test edilmiştir ve reaktif olanları ile TMA boyamaları gerçekleştirilmiştir. Bu ön antikor testleri her ne kadar projenin sürecini uzatıyor gibi görünse de TMA camlarının çok pahalı olması (2000-1000TL) ve çalışmayacak bir antikor için harcanmaması açısından önemlidir. Satın alınan bazı antikorlar IHC klinik koşullarda reaktivite göstermemiştir.

**Tablo 4. Doku boyamaları için planlanmış olan ve test aşamasındaki antikorlar**

<b>Proje kapsamında planlanan boyamalar</b>		
<b>Akciğer</b>	<b>Karaciğer</b>	<b>Meme</b>
HxE, CD133, ALDH1, CD166, CD44	HxE, CD133, CD44, EpCAM, OV6, CD13	HxE, CD133, ALDH, CD44, CD166
<b>Toplam görüntü alınacak doku parçacığı</b>		
<b>1050</b>	<b>1644</b>	<b>1130</b>

(Görüntüler 4-20X büyütme ile alınmıştır)

Prof. Dr. Ayşegül Üner danışmanlığında, kanserli dokulara yönelik değişik antikorların konsantrasyonlarını (1:50, 1:100, 1:200) değişik inkübasyon süreleri ile titrasyon deneyleri meme dokuları için uygun antikor konsantrasyonunu belirlenmesini yapıldı. Daha sonra meme TMA dizinleri yaygın olarak kullanılan KKH işaretleyicileriyle boyanmaya başlanmıştır (Şekil 1). Bu sayede boyanmış kanser dokularında kök hücre/tümör hücre oranları belirlenmeye başlanmıştır. Tablo 2'de belirtilen örnekler yaygın olarak kullanılan KKH işaretleyicisi antikor ile IHC boyandıktan sonra ışık mikroskobu yardımıyla 4X-10X-20X ve 40X olarak incelenmiş ve görüntüleri alınmıştır. Bu görüntüler Prof. Dr. Ayşegül Üner tarafında projede kullanılmaya uygun bulunmuştur. Karaciğer akciğer örneklerinde olduğu gibi veri setleri oluşturulması sürmektedir.



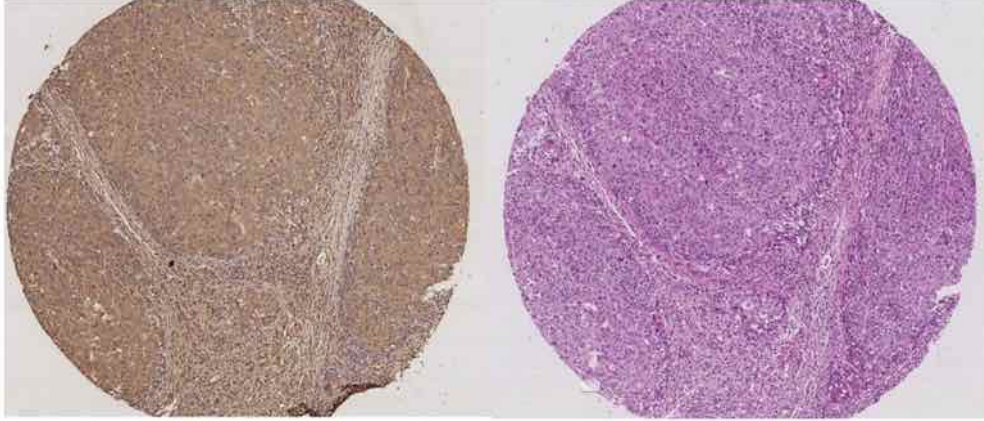
**Şekil 1. Seçilen örnek 3 meme TMA seri doku kesitlerinde EpCAM IHC boyaması gösterilmiştir. Camlar ışık mikroskobu altında 4X-10X-20X ve 40X büyütmelede incelenmiş ve görüntüleri alınmıştır.**

### **3.1.1 Dokuların görüntülerinin seçimi**

İmmunohistokimya ve H&E boyalı TMA camların görüntüleri proje danışmanımız Prof. Dr. Ayşegül Üner rehberliğinde incelenmiş ve görüntü analizi için uygun olanları seçilmiştir. Daha sonra bu camlar mikroskop altında tekrar görüntülenmiş ve Yaklaşık 1000 görüntü alınmıştır. Mikroskop altında manuel olarak olarak seçilen bu görüntüler yazılım geliştirilmesi sistemi için kullanılmak üzere etiketlenmesidir. Biyobelirteçlerle boyanmış imgeler yer gerçekliği etiketlerinin oluşturulması için kullanılmıştır, çünkü H&E boyanmış dokularda kanser kök hücreleri (KKH) çıplak gözle ayırt edilmemektedir.

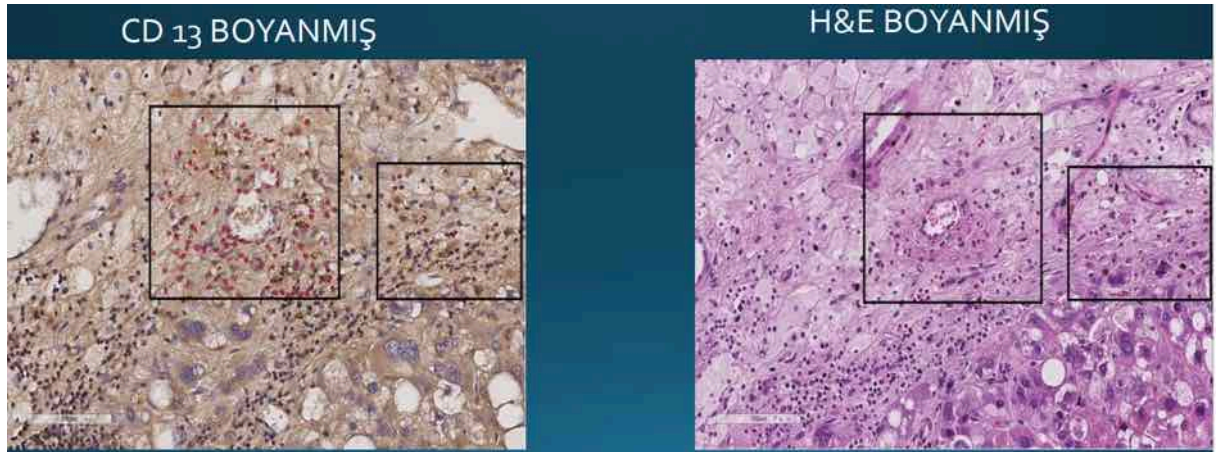
Histopatolojik imgelerin oluşturulması için öncelikle hastadan doku örneği alınır ve bu doku ince katmanlar halinde kesilir. Ardışık katmanlar birbirine görsel olarak çok benzese de tam

olarak aynı değildirler ancak boyanmış bir katman başka bir boyayla tekrar boyanamayacağı için bir katmanın hem CD13 hem de H&E boyanmış versiyonunu elde etmek mümkün değildir. Bu sebeple, ardışık doku katmanları CD13 ve H&E ile boyanarak neredeyse aynı görsel özelliklere sahip oldukları varsayılmıştır. Seri kesim CD13 ve H&E ile boyanmış ardışık iki katman Şekil 2'de görülmektedir. İki katman görsel olarak hemen hemen aynı doku yapısına sahiptirler. Bu varsayıma dayanarak KKH yüzdeleri CD13 imgeleri kullanılarak hesaplanmış ve karşılık gelen H&E imgeleri için kesin referans etiketleri oluşturulmuştur.



**Şekil 2. Aynı hastadan alınıp CD13 (solda) ve H&E (sağda) boyanmış seri kesim ardışık karaciğer doku katman imgeleri.**

Şekil 2 de örneği verilen İmmunohistokimya görüntüleri yarıotomatik olarak tek tek elle işaretlenmiştir. Gerçek veri seti oluşturulmuştur. 454 cam görüntüsü içeren bir set üzerinde yazılım geliştirilmiştir. Manuel etiketlenmiş görüntü örneği Şekil 2 de verilmiştir.

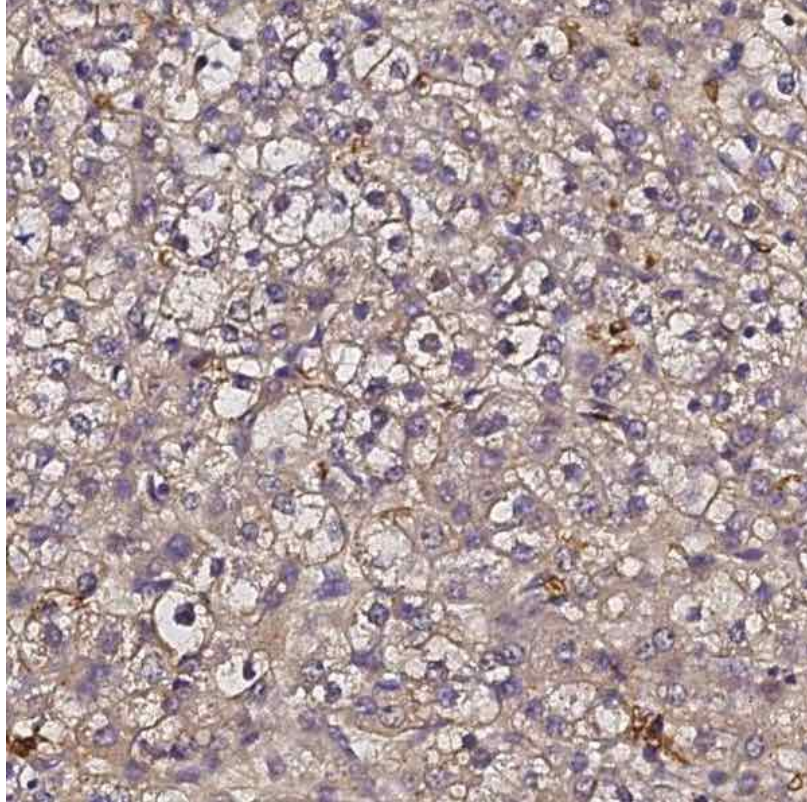


**Şekil 3. Solda manuel etiketlenmiş immunohistokimya görüntü örneği. Sağda o bölgeye denk gelen H&E görüntüsü verilmektedir.**

## 3.2 Metotlar

### 3.2.1 İmmünohistokimya İmgelerinde Kanser Kök Hücrelerinin Bulunması

Yukarda da belirtildiği üzere bir immünohistokimya boyası olan CD13 ile boyanmış kanserli dokularda kanser kök hücrelerinin (KKH) çekirdeklerinin kahverengiye boyandığı fark edilmiştir. Diğer çekirdekler 4. Şekilde de görüleceği üzere lacivert renge boyanmaktadır. Kahverengiye boyalı başka çekirdeklerde olabilir ama kahverengi çekirdekler sayılarak kök hücrelerin diğer hücrelere oranı tahmin edilebilir (Yamashita v.d. 2013). Kahverengi renge boyalı hücre çekirdeği sayma işi insan gözüyle yapılması zaman alan bir işlem olmasının yanı sıra insan kaynaklı hatalara da açıktır. Bu kısımda zamandan kazanmak ve insan kaynaklı hataları en aza indirmek amacıyla CD13 ile boyanmış imgelerdeki kahverengiye boyalı kanser kök hücrelerinin otomatik olarak bulunması amacıyla geliştirdiğimiz algoritma ve yazılım anlatılmaktadır. Yapay sinir ağı yapısına benzeyen kullanıcıyla etkileşimli bir yazılım geliştirilmiştir. Yazılım kanser kök hücrelerini ilk etapta doktora göstermekte daha sonra da doktorun düzeltmesine izin vermektedir. Doktorun belli bir bölgede yaptığı düzeltmelere göre resmin başka kısımlarında yapay sinir ağının ağırlıkları değiştirilerek otomatik olarak düzeltilme yapılmaktadır.



**Şekil 4 Kanserli karaciğer dokusuna ait kanser kök hücreleri içeren bir CD13 imgesi**



Şekil 4'te CD13 ile boyanmış kanserli bir karaciğer dokusunun resmi vardır. Çekirdekler genelde mavi renge boyanmıştır ama bazı çekirdekler koyu kahve renklidir. Bu gözlemden yola çıkılarak kanser kök hücrelerini çeşitli bölge betimleyicileri ile tanımlayabiliriz. Geliştirdiğimiz sistemde kanser kök hücrelerinin bulunmasında dört farklı bölge betimleyicisi kullanılmıştır:

1. Bölge Kovaryans Matrisi ,
2. Bölge Ortak Fark Matrisi,
3. Bölgenin Kırmızı-Yeşil-Mavi (RGB) Ortalama Değerleri,
4. Bölge "YCbCr", "HSV" renk uzaylarındaki ortalama Değerleri ve Kosinüs Açısı Ölçüsü

Aşağıda önce 1. ve 2. bölge betimleyicileri anlatılmaktadır. CD13 ile boyanmış imge önce küçük bölgelere ayrılmakta ve her bölgede kovaryans ve ortak fark matrisleri hesaplanmaktadır. Örüntü tanıma problemlerinde başarılı sonuçlar veren bölge kovaryans matrisi (Tuzel vd., 2006)'da aşağıdaki gibi tanımlanmıştır:

$$C_v = \frac{1}{N-1} \sum_{k=1}^N (f_k - \mu) \cdot (f_k - \mu)^T \quad (1)$$

1 numaralı denklemde  $f_k$ , k'ninci pikselin özellik vektörüdür. N seçilen bölge içinde bulunan piksel sayısını,  $\mu$  ise ortalama özellik vektörünü temsil etmektedir. Çalışmamızda kullanılan özellik vektörü  $f_k$  ve ortalama vektörü  $\mu$  aşağıdaki tanımlanmıştır:

$$f_k = \left[ R, G, B, \frac{dR(x,y)}{dx}, \frac{dR(x,y)}{dy}, \frac{d^2R(x,y)}{dx^2}, \frac{d^2R(x,y)}{dy^2} \right] \quad (2)$$

$$\mu = \left[ 0, 0, 0, \mu\left(\frac{dR(x,y)}{dx}\right), \mu\left(\frac{dR(x,y)}{dy}\right), \mu\left(\frac{d^2R(x,y)}{dx^2}\right), \mu\left(\frac{d^2R(x,y)}{dy^2}\right) \right] \quad (3)$$

yukarıdaki vektörlerde R,G,B girdileri sırasıyla piksele ait kırmızı, yeşil ve mavi renk bilgilerini temsil eder. Bu renk bilgilerinin yanı sıra  $f_k$  vektörüne, kırmızı renk kanalında bulunan piksellerin x ve y eksenlerindeki birinci ve ikinci türev bilgileri de eklenmiştir.  $\mu$  vektöründe renk bilgilerinin ortalama değerleri 0 alınarak renk bilgisinin korunması amaçlanmış ve deneysel olarak daha iyi sonuçlar elde edildiği gözlemlenmiştir. Ortak fark matrisi ise (4) numaralı denklemdeki gibi tanımlanmıştır.

$$C_d = \frac{1}{N-1} \sum_{k=1}^N (f_k - \mu) \oplus (f_k - \mu)^T \quad (4)$$

Yukarıdaki denklemde  $\oplus$  operatörü çarpma işlemi yerine kullanılmaktadır ve iki sayı, a ve b, için aşağıdaki şekilde tanımlanır:

$$a \oplus b = \text{sign}(a \times b)(|a| + |b|) \quad (5)$$

6 numaralı denklemde gösterilen  $7 \times 7$  ebatındaki  $C_{ROI}$  matrisi, analiz edilmesi istenen  $11 \times 11$ 'lik bir bölge için hesaplanan bir kovaryans matrisini göstermektedir:

$$C_{ROI} = \begin{bmatrix} C_{1,1} & \cdots & C_{1,7} \\ \vdots & \ddots & \vdots \\ C_{7,1} & \cdots & C_{7,7} \end{bmatrix} \quad (6)$$

$C_{ROI}$  matrisi simetriktir ve bu özellik kullanılarak,  $11 \times 11$  piksel boyutundaki bir bölge,  $C_{ROI}$  matrisinin alt üçgen değerleri ile temsil edilebilmektedir. Bu alt üçgen değerleri daha sonra bölgenin öznitelik vektörü olarak kullanılmıştır:

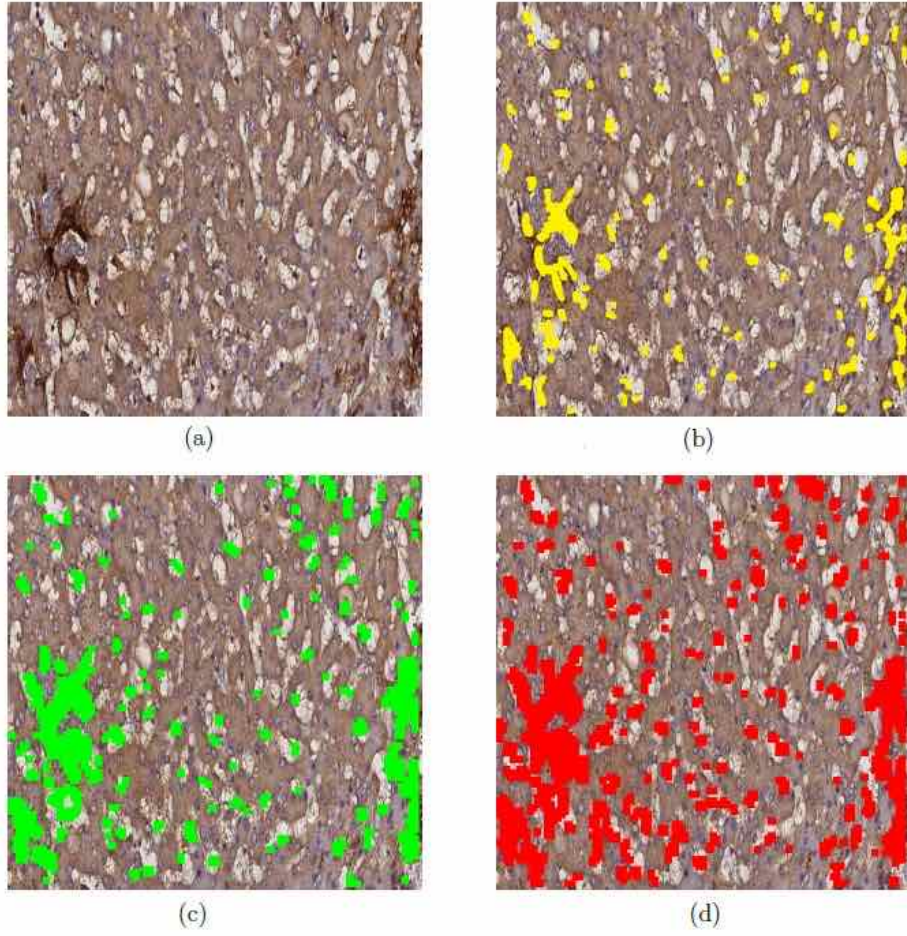
$$z_k = [C_{1,1}, C_{2,1}, C_{2,2}, C_{3,1}, \dots \dots C_{7,7}] \quad (7)$$

Ortak fark matrisi ( $C_d$ ) kullanıldığında ise (6) numaralı denklem ortak fark değerleri ile doldurulmaktadır.

Bölge kovaryans matrisi ve bölge ortak fark matrislerinin CD13 imgelerindeki kanser kök hücrelerinin bulunmasına uygulanmıştır. Bu öznitelik vektörleri ile destek vektör makinaları (support vector machine) ve yapay sinir ağları eğitilmiştir. 5. Şekilde destek vektör makinası ile elde edilen bir örnek gösterilmektedir. En iyi sonuçlar ortak fark matrisi kullanıldığında elde edilmektedir.

Ayrıca  $11 \times 11$  piksellik bölgelerin averaj kırmızı-yeşil ve mavi (RGB), ve HSV ve YCbCr renk uzaylarındaki ortalama değerleri çıkarılmış ve bunlarla da ek öznitelik vektörleri olarak kullanılmıştır. Bu öznitelik vektörleri ile de destek vektör makinaları ve yapay sinir ağları eğitilmiştir.

Dört betimleyici uyarlamalı bir algoritma ile karıştırıldığında daha da iyi sonuçlar elde edilebilmektedir.



**Şekil 5 CD13 ile boyanmış bir kanserli imgenin (a), kahverengi bölgelerin işaretlendiği yer gerçekliği imgesi (b), ortak fark metodu (c), ve kovaryans metodu (d) ile elde edilen sonuçlar. En iyi sonuçlar ortak fark matrisi kullanıldığında elde edilmektedir.**

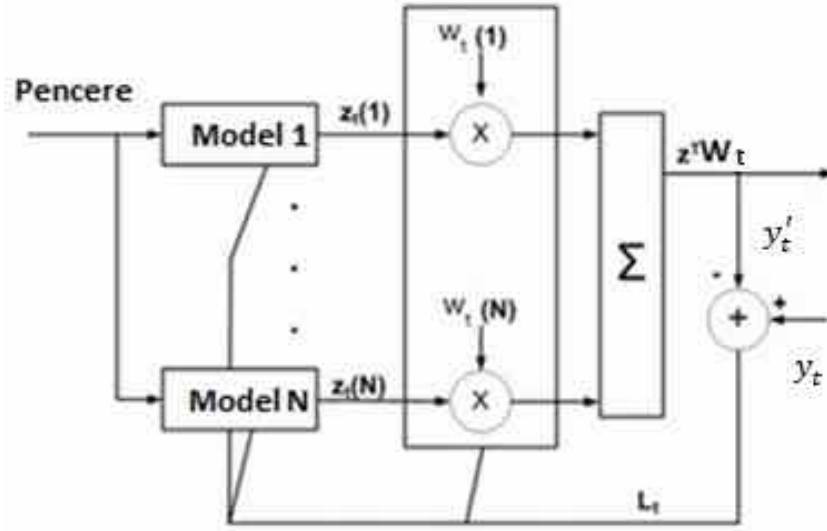
### **3.2.2 Betimleyici Karıştırma Metodu (BKM) ve CANSTEM yazılımı**

Betimleyici karıştırma metodunun amacı farklı öznitelik vektörleri ve farklı makine öğrenme yöntemleri ile eğitilmiş bölge betimleyicilerinin sonuçlarını birleştirerek KKH'lerinin bulunmasındaki başarıyı yükseltmektir. BKM ayrıca patologdan gelen düzeltmelere göre kendi tahmin ettiği ağırlık değerlerini de değiştirmektedir. 6. Şekilde BKM'nin ana hatları gösterilmektedir.

Yukarda da belirtildiği üzere dört bölge betimleyici yöntemin her biri için bir destek vektör makinası ve bir yapay sinir ağı eğitilmiştir. 6. Şekilde N tane sınıflandırıcı olduğu varsayılmıştır. Bizim durumumuzda  $N=8$ 'dir. İlk aşamada kullanıcıya destek vektör makinası (DVM) ve yapay sinir ağlarının (YSA) sonuçları eşit ağırlandırılarak gösterilmektedir. Daha sonra da kullanıcının geri beslemesine göre ağırlıklar "en küçük kare averaj" (LMS-Least

Mean Square) benzeri bir algoritma ile değiştirilmektedir.

Algoritma gerçek zamanlı öğrenmeye açık bir algoritmadır ve gerçek zamanlı öğrenmenin kolayca uygulanabilmesi için bir grafik ara yüzü (GUI-Graphical User Interface) MATLAB yazılımı kullanılarak hazırlanmıştır. Algoritmanın çalışma prensibi şöyle özetlenebilir: Kullanıcı tarafından yazılıma yüklenen CD13 imgesi 11x11 piksel boyutundaki kayan bir pencere tarafından taranır. Her küçük pencere için destek vektör ve yapay sinir ağı sonuçları elde edilir.



**Şekil 6 Eğitilen modellerin işlenen pencere üzerine verdikleri kararların LMS tabanlı algoritma ile birleştirilmesi**

İlk aşamada ağırlıklar  $w_i = \frac{1}{N}$  olarak varsayılır ve sonuç pozitif ise kullanıcıya gösterilir. 6. Şekilde bir bölge için kullanıcının kararı  $y_t$  ile gösterilmiştir. Kullanıcı bu kararı değiştirmek isterse ekrandaki ilgili bölgeye tıkladığında pozitif (+1) veya negatif (-1) kararına göre  $y_t$  değeri oluşturulur. 6. Şekildeki  $w_{t+1} = [w_{t+1,1} \dots w_{t+1,N}]^T$  yenilenmiş katsayıları temsil ederken,  $w_t = [w_{t,1} \dots w_{t,N}]^T$  yenilenmemiş katsayı vektörünü,  $D_t = [z_t(1), \dots, z_t(N)]^T$  ise modellerin karar değerlerini temsil etmektedir. Bir sonraki değerler aşağıdaki şekilde hesaplanır:

$$w_{t+1} = w_t + \frac{L_t}{\|D_t\|^2} D_t \quad 8$$

Yukardaki denklemdaki hata değeri  $L_t$  de aşağıdaki şekilde hesaplanır :

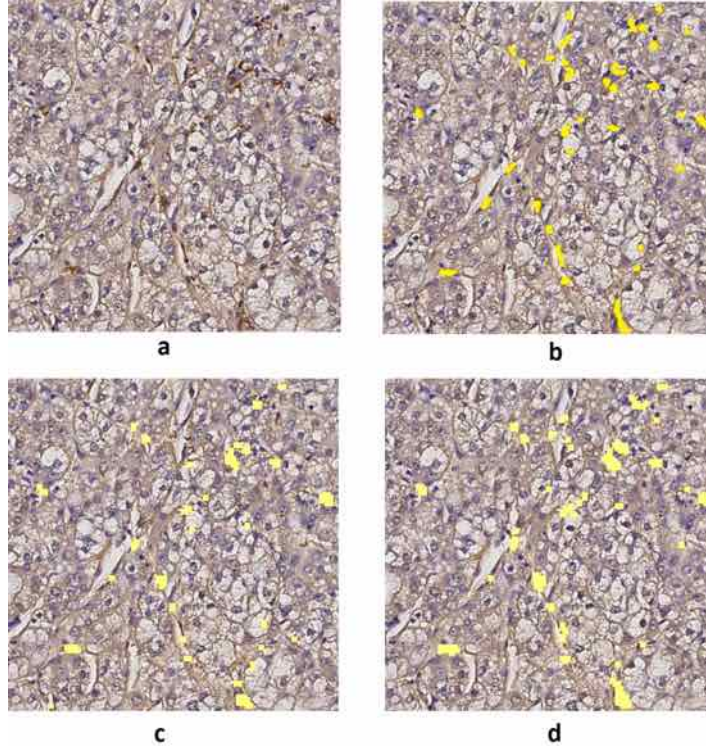
$$L_t = y_t - y_t' \quad 9$$

buradaki  $y_t$  kullanıcı tarafından girilen bölge sınıfı değeri iken,  $y_t' = z^T w_t$ , 11x11'lik imge bloğunun N=8 model tarafından ağırlıklandırılarak tahmin edilmiş değeridir. (9). Denklemdaki yaklaşım meşhur uyarılma algoritması LMS yapısına dayalıdır.

İmgenin başlangıç katsayıları ile elde edilen sonucu patoloğa geliştirilen grafik arayüzü vasıtasıyla gösterilir. Patolog imgedeki düzeltmeleri yine arayüz aracılığı ile yapar. Gelen yeni bilgiler doğrultusunda LMS tabanlı algoritma ile katsayıları yeniler ve patoloğun istediği sonuçlar elde edilmeye çalışılır. İstenilen sonuç bir kaç denemeden sonra makine öğrenmesi ile elde edilemezse patoloğun istediği bölgenin ortak fark matrisi hesaplanıp imgedeki diğer bölgeler ile uzaklığı hesaplanır. Bu yazılımda iki matris  $(C_1, C_2)$  arasındaki uzaklık aşağıdaki gibi tanımlanmıştır:

$$d(C_1, C_2) = \sum_{i=1}^P \left[ \sum_{j=1}^P \frac{|C_1(i, j) - C_2(i, j)|}{(C_1(i, i) + C_2(i, i))} \right] \quad 10$$

İki bölgenin bir birlerine benzer olup olmadıklarını belli bir eşik değeri kullanarak belirlemek yerine bir "sigmoid" fonksiyonu kullandık. Bu sayede herhangi bir eşik değeri kullanılmamıştır. Katsayılar ve uzaklık ölçütleri kayıt edilmektedir. Böylece katsayılar benzer imgelerde kullanılabilir ve her imgede katsayı hesabı yapılması gerekmez.



**Şekil 7 CD13 ile boyanmış bir kanserli imgenin (a), yer gerçekliği imgesi (b), ilk sonuç (c), ve betimleyici karıştırma metoduyla eğitilmiş katsayılarla elde edilen sonuç (d).**

Geliştirdiğimiz yazılıma CANSTEM yazılımı adını verdik. Deneysel sonuçlar bulgular kısmında verilmiştir.

### 3.2.3 CANSTEM Yazılımın Kullanıcı Arayüzü

Geliştirilen yazılımın kullanıcı ile etkileşimi için bir arayüz hazırlanmıştır. Bu arayüzde bulunan özellikler sayesinde kullanıcı geri dönüşlerini algoritmaya verebilmektedir.

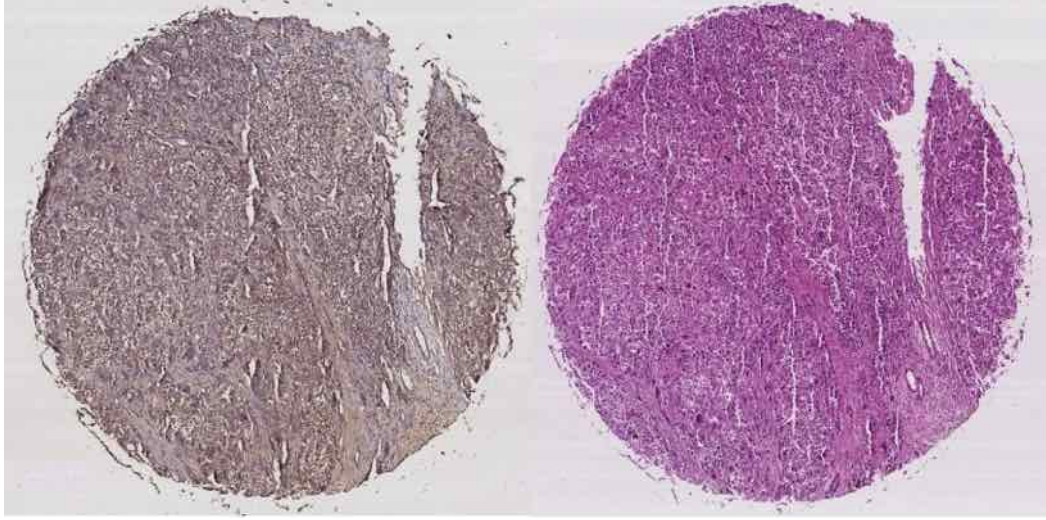
Geliştirilen yazılımın arayüzünde kullanıcı önce içindeki KKH'lerin bulunmasını istediği imgeyi "Upload Image (İmge Yükle)" tuşu ile yükler ve algoritmanın çalışması için "Find CSC (KKH Bul)" tuşuna basar (EK1). Yazılım ilk ağırlık vektörünü kullanarak KKH'leri bulur ve kullanıcıya ilk sonucu gösterir. Kullanıcı bu ilk sonuç üstünde işaretlemeler yaparak algoritmaya düzeltilmesi gereken yerlerin bilgilerini verebilir. Arayüzde bulunan "CSC + (KKH +)" tuşu, kullanıcının yazılımın işaretlediği KKH'leri işaretlemesini sağlarken, "CSC- (KKH -)" tuşu ise KKH olmadığı halde pozitif olarak işaretlenmiş yerleri belirtmek amacı ile kullanılır. Kullanıcı yaptığı işaretlemeleri "Undo (Geri Al)" tuşu ile geri alabilir. Uygun gördüğü işaretlemeler bittiğinde "Update Weights (Ağırlıkları Güncelle)" tuşuna basarak, yaptığı düzeltmelere uygun olarak yazılımın çalışmasını sağlar. Yazılım kullanıcıdan gelen bilgilere göre gerekli değişiklikleri yapar ve yeni sonucu tekrar kullanıcıya gösterir. Bu işlem kullanıcı çıkan sonuçtan memnun kalana kadar devam ettirilebilir ve kullanılan ağırlık vektörü her seferinde kayıt edilir. Bu kayıt edilen ağırlık vektörü, daha sonra başka imgelerdeki KKH'lerin bulunmasında kullanılır. Böylece her imge için ağırlık vektörünü eğitmek gerekmez ve zamandan büyük ölçüde tasarruf edilmiş olunur. Bir süre sonra ağırlık vektörü oldukça iyi eğitilmiş olacağından dolayı kullanıcının artık ağırlık vektörlerini güncellemesine gerek kalmayacaktır. Kullanıcı EK1'de detaylı anlatıldığı sekliyle "CSC+ (KKH+)" tuşu ile imgenin sağ yarısında iki KKH işaretlemiş ve ağırlıkları tekrar güncellemiştir. Güncellenen ağırlıklar programın bir sonraki turda işaretli bölgeleri tespit etmesini sağlamıştır. İkinci turda tespit edilen yeni KKH'leri kırmızı çember içerisinde görülebilmektedir.

### 3.2.4 H&E Boyanmış Dokularında Histopatolojik Kanser Kök Hücre İmge Sınıflandırılması

H&E boyası düşük maliyetli bir histopatolojik boya tipidir. CD13 boyanmış imgelerin aksine H&E boyanmış imgelerde Kanser Kök Hücrelerini (KKH) çıplak gözle ayırt mümkün değildir. Ayrıca rutin tanı ve takip te doku kesitleri öncelikle H&E ile boyanır ve eğer gerekiyorsa IHK boyaması yapılır. Ayrıca rutinde tümör doku heterojenitesi (KKH) için IHK de dahil olmak üzere bir yaklaşım bulunmamaktadır. Bu sebeple H&E imgelerini KKH yüzdelerine göre sınıflandırabilecek bir yöntem daha ulaşılabilir ve ekonomik olacaktır. Bu bölümde, H&E imgelerini KKH yüzdelerine göre sınıflandıran bir örüntü tanıma algoritması geliştirilmiştir.

Histopatolojik imgelerin oluşturulması için öncelikle hastadan doku örneği alınır ve bu doku

ince katmanlar halinde kesilir. Ardışık katmanlar birbirine görsel olarak çok benzese de tam olarak aynı değildirler ancak boyanmış bir katman başka bir boyayla tekrar boyanamayacağı için bir katmanın hem CD13 hem de H&E boyanmış versiyonunu elde etmek mümkün değildir. Bu sebeple, ardışık doku katmanları CD13 ve H&E ile boyanarak neredeyse aynı görsel özelliklere sahip oldukları varsayılmıştır. Ardışık CD13 ve H&E ile boyanmış ardışık iki katman Şekil 2'de gösterilmiştir. Şekil 8'de başka bir hastadan alınan iki katmanın görüntüleri vardır. İki katman görsel olarak hemen hemen aynı doku özelliklerine sahiptir. Bu varsayıma dayanarak KKH yüzdeleri CD13 imgeleri kullanılarak hesaplanmış ve karşılık gelen H&E imgeleri için kesin referans etiketleri oluşturulmuştur.



**Şekil 8 Aynı hastadan alınıp CD13 (solda) ve H&E (sağda) boyanmış ardışık karaciğer doku katman imgeleri.**

Yukarıda da belirtildiği gibi kanser kök hücreleri H&E ile boyanmış imgelerde diğer hücrelerden göz ile ayrıştırılamıyor. Proje ara raporlarında da aktardığımız üzere CD13 ile boyanmış imgelerde bulunan kanser kök hücrelerinin etraflarında boşluklar oluşturduğu gözlemlenmiştir. Bu boşlukların H&E imgeleri üzerinde örüntü (texture) farkı oluşturduğunu gözlemledik. Bu özellikleri kullanarak örüntü tanıma ile karaciğer kanser kök hücrelerinin oranlarını H&E görüntüleri üzerinden yaklaşık olarak belirleyebildik. Ayrıca bazı hücrelerin diğerlerine göre daha koyu boyandığını gözlemledik. Bu daha koyu olanların kanser kök hücre olma ihtimali olduğuna inanıyoruz.

CD13 ile boyanmış resimlerde de olduğu gibi örüntü tanıma problemlerinde başarılı sonuçlar veren bölge kovaryans matrisi [Porikli,2006], ortak fark matrisi gibi betimleyicileri H&E görüntülerinin sınıflandırılmasında kullanıldı.

Bu çalışmada kovaryans matrisini oluştururken çeşitli  $f_k$  vektörleri denenmiştir. Bunlardan birincisi aşağıdaki gibidir:

$$f_{k1} = [R, G, B, L, a, b, H, E] \quad 11$$

yukarıdaki denklemde R,G,B değerleri piksele ait kırmızı, yeşil ve mavi renk uzayındaki bilgisini temsil ederken, L,a,b ve H,E ; L,a,b ve H & E renk uzaylarındaki bilgilerini temsil eder. H&E renk uzayı (Cosatto vd., 2008)'de tanımlanmış ve H&E imgelerinde kullanılmak üzere geliştirilmiştir (Suhre vd., 2013). H&E boyası ile boyanmış imge CMY renk düzlemine aktarılmış ve CMY renk uzayında yapılan izdüşümler sonrası her piksel için H ve E değerleri aşağıdaki adımlar kullanılarak elde edilir. Önce  $\gamma$  ve  $\epsilon$  değerleri CMY vektöründen elde edilir:

$$\gamma = \frac{\sum w_i P_i}{\sum w_i}, \quad w_i = (P_i^{CMY})^4 \quad 12$$

ve

$$\epsilon = \frac{\sum a_i P_i}{\sum a_i}, \quad a_i = \left| P_i - \frac{(P_i \cdot \gamma) \gamma^4}{\gamma^2} \right| \quad 13$$

yukarıda  $P_i$  pikselin CMY renk uzayındaki renk vektörünü temsil etmektedir. Daha sonra da izdüşüm yapılarak H ve E değerleri elde edilir:

$$H = \gamma - \frac{(\gamma \cdot \epsilon) \epsilon}{|\epsilon|^2}, \quad E = \epsilon - \frac{(\gamma \cdot \epsilon) \gamma}{|\gamma|^2}, \quad 14$$

İkinci özellik vektörü olarak da H&E ile boyanmış imgelerin pembemsi renkte olmasından dolayı kırmızı renk kanalındaki türevler vektöre eklenmiştir. Pikselin yatay (x) ve dikey (y) eksenlerindeki birinci ve ikinci türevleri kırmızı renk kanalı üzerinden hesaplanarak aşağıdaki vektör oluşturuldu:

$$f_{k2} = \left[ R, G, B, L, a, b, \frac{dR(x,y)}{dx}, \frac{dR(x,y)}{dy}, \frac{d^2R(x,y)}{dx^2}, \frac{d^2R(x,y)}{dy^2}, H, E \right] \quad 15$$

Üçüncü özellik vektörümüz ise kontrol amaçlı olarak sadece renk değerleri



kullanılarak aşağıdaki şekilde oluşturuldu:

$$f_{k3} = [R, G, B, H, E] \quad 16$$

Son olarak da çok elemanlı büyük bir vektör de denendi:

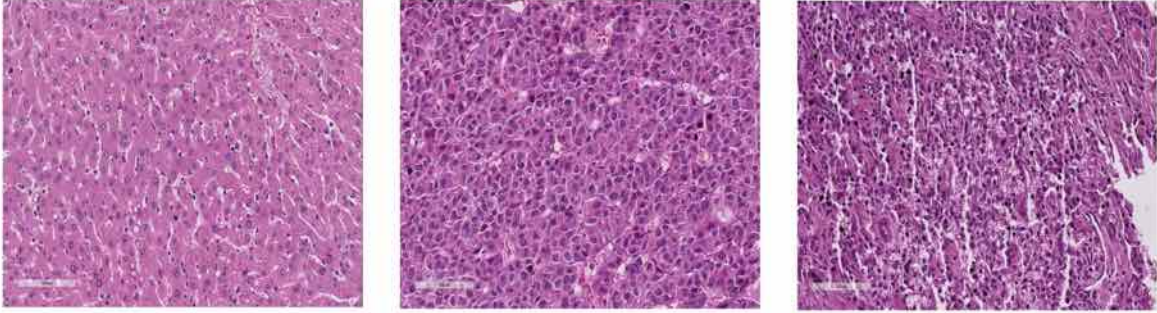
$$F = [hist(R) hist(R_t) hist(G) hist(G_t) hist(B) hist(B_t) hist(Y) hist(Y)' hist(Y_t) hist(U) hist(U)' hist(U)'' hist(V) hist(V)' hist(V)'' hist(H) hist(H_t) hist(E) hist(S) hist(S_t)]^T \quad (17)$$

Bu vektörde  $hist(R)$ , RGB renk uzayından “R” kırmızı renk kanalının histogramını, “G” yeşil renk kanalını, “B” ise mavi renk kanalını ifade eder. HE renk uzayından “H” Hematoxylin renk kanalını, “E” Eosin renk kanalını ifade eder. YUV renk uzayından “Y” parlaklık kanalını, “U” ve “V” kanalları renklilik kanallarını ifade eder. HSV renk uzayından “S” ise doygunluk kanalını ifade eder.  $hist(\cdot)$  fonksiyonu girdi olarak verilen tek kanal imgenin 128 elemanlı histogramını hesaplar.  $hist(\cdot)'$  fonksiyonu  $hist(\cdot)$  fonksiyonunun birinci dereceden kesikli türevini ifade eder.  $hist(\cdot)''$  fonksiyonu ise  $hist(\cdot)$  fonksiyonunun ikinci dereceden kesikli türevini ifade eder.  $(\cdot)_t$  operatörü belirtilen kanalda piksel eşik değerlemesini ifade eder. Piksel eşik değerlemesinin amacı koyu renkli bölgeleri daha iyi temsil eden öznelikler çıkarmaktır. Deneylerde deneysel verilere dayanarak eşik değeri 200 olarak seçilmiştir. R, G, B, Y, U, V ve E kanallarında piksel değeri 200'den büyük olan pikseller yok sayılmıştır. H ve S kanallarında ise piksel değeri 200'den küçük olan pikseller yok sayılmıştır. Bu şekilde hücrel bölgelere daha anlamlı öznelikler çıkarılması sağlanmıştır. Yukarıdaki denklemde  $hist(\cdot)$  fonksiyonu  $128 \times 1$  boyutunda ve  $[0, 255]$  aralığında her iki değer için bir değer elde edilmiştir. Farklı kanallardan elde edilen histogram çıktıları (öznelik vektörleri) yan yana eklenerek her imge için  $20 \times 128$  boyutunda bir öznelik matrisi  $F$  üretilir.

CD13 ile boyalı resimlerde olduğu gibi kovaryans matrisinden farklı olarak ortak fark matrisi (Tuna vd., 2009) yaklaşımı da denemiştir. Sınıflandırma sırasında bölge kovaryans  $C_v$  ve bölge ortak fark  $C_d$  matrisleri hesaplandıktan sonra matrislerin simetri özelliğinden yararlanarak alt üçgen değerlerinden bir vektör oluşturulmuştur. Destek vektör makinalarını alt üçgen değerleri ve matrislerin öz değerleri ile eğittik. Bu vektörler elimizdeki 200 imgenin tümü için hesaplandı.

Yukarıda belirtildiği üzere CD13 ile boyanmış karaciğer imgelerinden kanser kök hücrelerini sayarak aynı hastaların H&E imgeleri üzerinde 3 sınıf oluşturduk. 1. Sınıf kansersiz karaciğer imgelerinden oluşturuldu. 2. sınıf görüntüler kök hücrelerinin diğer hücrelere oranı %5'den az olan görüntülerden oluşturulmuştur; 3. sınıf ise %5'in üstünde kök hücre olan H&E resimlerinden oluşturulmuştur. Aslında bu oranları CD13'le boyanmış

imgelerdeki kahverengi boyalı çekirdekleri sayarak belirledik. Bu hastalardaki kök hücre oranları daha düşük olabilir.



**Şekil 9: Sırasıyla Sınıf-1,Sınıf-2 ve Sınıf-3'e ait H&E ile boyanmış imgeler (20x büyütme).**

Şekil 9'da H&E ile boyanmış Sınıf-1, Sınıf-2 ve Sınıf-3 imgelerinden örnekler vardır. Bu üç sınıf imgeyi çeşitli sınıflandırma yöntemleri ile sınıflandırdık. CANSTEM yazılımında başarılı sonuçlar elde ettiğimize inanıyoruz.

### **3.2.5 H&E ile Boyanmış Akciğer İmgelerinin Sınıflandırılması**

Akciğer imgelerinin analiz metodu karaciğer imgelerine çok benzerdir. Aradaki tek fark akciğer imgelerinde yer gerçekliği oluşturmak için CD166 IHK boyasının kullanılmasıdır. Akciğer imgelerinden kesilen ardışık doku katmanları CD166 ve H&E boyaları ile boyanır. Akciğer imgelerinde de karaciğer imgelerine benzer şekilde kesin referans etiketleri CD166 boyanmış imgelerden elde edildi. Akciğer H&E imgeleri karaciğer imgelerinde olduğu gibi sınıflandırıldı.

## 4. BULGULAR

### 4.1 H&E ile Boyanmış Karaciğer İmgelerinin Sınıflandırılma Sonuçları

Bir önceki bölümdeki öz nitelik vektörlerini kullanarak elde ettiğimiz sonuçları aşağıda tablolara koyduk. Bu tablolardan da görüleceği üzere bu üç sınıf üzerinde yüksek tanıma oranlarına ulaşılmıştır. Karaciğer imge kümemiz 3 sınıfa ait karaciğer doku imgelerini içerir. Derece I (17 hastadan alınan 119 imge), Derece II (17 hastadan alınan 151 imge) ve normal (9 hastadan alınan 184 imge) doku imgeleri. Eğitim sırasında destek vektör makinalarına imgelerin %60'ı model eğitim için verilmiştir. Kalan %40'ın yarısı katlamalı çapraz doğrulama için diğer yarısı ise eğitilen modellerin test edilmesi amacıyla kullanılmıştır. Destek vektör makinesinin çekirdek fonksiyonu olarak çembersel taban fonksiyonu (radial basis function) seçilmiş ve bu fonksiyonun en iyi C ve Y değerleri 10-katlamalı çapraz doğrulama ile bulunmuştur. Modellerin doğruluk oranlarının yanı sıra modelleri daha anlamlı bir şekilde yorumlayabilmek adına modellerin Genel Başarı, Keskinlik, Tanıma, F1 skoru (Yang ve Liu, 1999) ve Özgüllük (specificity) değerleri hesaplanmıştır:

$$\text{Genel Başarı} = \frac{TP + TN}{TP + TN + FP + FN} \quad 88$$

bu denklemdaki TP, TN, FP ve FN elemanları sırasıyla gerçek pozitif, gerçek negatif, yanlış kabul ve ikinci tür hatayı temsil etmektedir. Özgüllük, Tanıma, Keskinlik ve F1 tanımları da aşağıdaki gibidir:

$$\text{Özgüllük} = \frac{TN}{TN + FP} \quad 19$$

$$\text{Tanıma} = \frac{TP}{TP + FN} \quad 20$$

$$\text{Keskinlik} = \frac{TP}{TP + FP} \quad 21$$

$$F1 = 2 * \frac{\text{Keskinlik} * \text{Tanıma}}{(\text{Keskinlik} + \text{Tanıma})} \quad 22$$

F1 sonucu keskinlik ve tanımanın harmonik ortalamasıdır. F1 sonucu [0,1] arası değişir.

**Tablo 5: Kovaryans matrisine dayalı metotta çeşitli öznitelik vektörleri kullanılarak elde edilen sonuçlar**

Kovaryans Metodu Sonuçları					
	Genel Başarı(%)	Özgüllük	Tanıma	Keskinlik	F1
$f_{k1}$	95.00	0.96	0.93	0.93	0.95
$f_{k1} - \text{özdeğer}$	87.50	0.90	0.82	0.82	0.82
$f_{k2}$	95.00	0.97	0.95	0.92	0.93
$f_{k3}$	75.00	0.75	0.91	0.61	0.73

**Tablo 6: Ortak fark matrisine dayalı metodun sonuçları**

Ortak Fark Metodu Sonuçları					
	Genel Başarı(%)	Özgüllük	Tanıma	Keskinlik	F1
$f_{k1}$	92.00	0.96	0.93	0.86	0.88
$f_{k1} - \text{özdeğer}$	90.00	0.94	0.87	0.90	0.88
$f_{k2}$	92.50	0.95	0.90	0.94	0.91
$f_{k3}$	90	0.92	0.96	0.81	0.87

En iyi sonuçlar  $f_{k1}$  öznitelik vektörü kullanıldığında kovaryans matrisinde elde edilmiştir. Sınıflandırma sonuçları kanımızca başarılıdır.

Bir sonraki kısımda da destek vektör makinaları yerine en yakın komşuluk sınıflandırıcısı ve SIFT vektörleri kullanarak sınıflandırma yaptık. Benzer sonuçlar elde edildi. Karaciğer kanserinde kök hücre oranı bulunmasında H&E resimlerinin kullanılabilineceğini gözlemledik.

*H&E ile Boyanmış Karaciğer İmgelerinin SIFT vektörlerinin Özyüz (Eigenface) Benzeri bir Metot ile Sınıflandırılması Sonuçları*

SIFT (ölçekten bağımsız özellik transformu) literatürde yaygın olarak kullanılan bir metottur. Biz H&E ile boyanmış görüntülere SIFT algoritmasını uyguladık.  $\phi_n$  SIFT algoritmasının çıkardığı 1x128 boyutlu bir özellik vektörü olsun. Hızlı karşılaştırma

yapabilmek için özyüz (eigenface) metoduna benzer bir yaklaşım geliştirdik (Turk ve Pentland, 1991). Önce bütün resimler için bir kovaryans matrisi oluşturduk.

$$C = \frac{1}{L \times M} \sum_n (\Phi_n - \mu)^T (\Phi_n - \mu) \quad 93$$

Eğitim kümemizden ortaya çıkan C matrisi 128x128'lik bir matristir. Bu matrisin özdeğer  $\lambda_k$  ve özvektörlerini  $u_k$  hesapladık. İki resmi karşılaştırmak için aşağıdaki adımları izledik:

- (i) Birinci resmin SIFT vektörleri  $\gamma_1, \gamma_2, \dots, \gamma_k$  olsun. İkinci resmin SIFT vektörleri de  $\zeta_1, \zeta_2, \dots, \zeta_k$  olsun.
- (ii) Birinci resmin SIFT vektörlerini  $u_k$  özvektörleri üzerine iz düşürüyoruz.  
 $v_k = u_k^T (\gamma_k - \mu), \quad k = 1, 2, \dots, 128$   
Aynı şekilde ikinci resmin SIFT vektörlerini de  $u_k$  özvektörleri üzerine iz düşürülmektedir.  
 $w_k = u_k^T (\zeta_k - \mu), \quad k = 1, 2, \dots, 128$
- (iii) Yukardaki v ve w vektörlerini kullanarak iki resim  $I_1$  ve  $I_2$  arasındaki uzaklık aşağıdaki şekilde hesaplanabilir.

$$d(I_1, I_2) = \|v - w\|^2 = \sum_k (v_k - w_k)^2 \quad (24)$$

Bir önceki kısımda da belirtildiği üzere H&E boyanmış karaciğer imgelerinden üç sınıf oluşturduk. Bunlar düşük kök hücre oranlı (%5'ten az), yüksek kök hücre oranlı (%5'ten fazla) ve kansersiz gruplardır. Bir test imgesinin bu üç gruptan hangisine ait olduğunu belirlemek için en yakın komşuluk (nearest neighbor - NN) sınıflandırıcısını kullandık. Deneysel çalışmalar sonucunda Tablo 6'te görüldüğü üzere %94.4 sınıflandırma başarısı elde ettik. 13 numaralı denklemde kovaryans matrisi yerine ortak fark matrisi de kullanılıncaya yine aynı sınıflandırma başarılarını elde ettik. Bu oranlar bir önceki bölümde anlatılan oranların biraz altındadır. Ancak, kök hücre oranlarının H&E imgelerinden örüntü analizi yapılarak bulunabileceğini göstermektedir.

Tablo 6 ve 7 elde edilirken 1. sınıfta 43 imge, 2. sınıfta 39 imge, 3. sınıfta ise 64 imge bulunmaktadır. Görüntü boyutları 300x300'dür. Tablo 7, 8 ve 9'da da komşuluk sayısı artırılarak elde edilen sonuçlar sunulmuştur. Tablo 11'de de ortak fark matrisine dayalı sonuçlar vardır.

**Tablo 7 3-NN Kullanarak Elde Edilen Sınıflandırma Başarı Oranı**

	Sınıf 1	Sınıf 2	Sınıf 3	Sınıf Keskinliği
Sınıf 1	42/43	0/43	1/43	97.7%
Sınıf 2	1/39	32/39	6/39	82.1%
Sınıf 3	0/64	5/64	59/64	92.2%
Sınıf Tanıması	97.7%	86.5%	89.4%	Genel Başarı 91.1%

**Tablo 8 5-NN kullanarak elde edilen sınıflandırma başarı oranı**

	Sınıf 1	Sınıf 2	Sınıf 3	Sınıf Keskinliği
Sınıf 1	41/43	1/43	1/43	95.3%
Sınıf 2	2/39	31/39	6/39	79.5%
Sınıf 3	0/64	1/64	63/64	98.4%
Sınıf Tanıması	95.3% 4.7%	93.9% 6.1%	90.0% 10.0%	Genel Başarı 92.5%

**Tablo 9 8-NN kullanarak elde edilen sınıflandırma başarı oranı**

	Sınıf 1	Sınıf 2	Sınıf 3	Sınıf Keskinliği
Sınıf 1	40/43	2/43	1/43	93.0%
Sınıf 2	2/39	29/39	8/39	74.4%
Sınıf 3	0/64	1/64	63/64	98.4%
Sınıf Tanıması	95.2%	81.3%	87.5%	Genel Başarı 90.4%

**Tablo 10 10-NN kullanarak elde edilen sınıflandırma başarı oranı**

	Sınıf 1	Sınıf 2	Sınıf 3	Sınıf Keskinliği
Sınıf 1	40/43	2/43	1/43	93.0%
Sınıf 2	2/39	31/39	6/39	79.5%
Sınıf 3	0/64	1/64	63/64	98.4%
Sınıf Tanıması	95.2%	91.2%	90.0%	Genel Başarı 91.8%

**Tablo 11 Ortak fark matrisi ve 3-NN kullanarak elde edilen sınıflandırma başarı oranı**

	Sınıf 1	Sınıf 2	Sınıf 3	Sınıf Keskinliği
Sınıf 1	42/43	0/43	1/43	97.7%
Sınıf 2	1/39	32/39	6/39	82.1%
Sınıf 3	0/64	5/64	59/64	92.2%
Sınıf Tanıması	97.7%	86.5%	89.4%	Genel Başarı 91.1%

#### *Ana Bileşen Analizi (PCA) Algoritması ile elde edilen sonuçlar*

Bu bölümde Ana Bileşen Analizi (Principal Component Analysis: PCA) algoritmasının değiştirilmiş bir versiyonu ile yaptığımız deneysel sonuçlar vardır. PCA algoritması geleneksel özyüz algoritmasından farklı olarak imgelerin piksel değerleri yerine imgelerin öznitelik matrislerinin ortak değişinti matrisleri değiştirilmiş PCA algoritmasına girdi olarak kullanır. Sınıflandırma sonuçları kovaryans matrisine dayalı metotlara göre daha düşük çıkmıştır. Ancak PCA algoritması kanserli dokuları ayırıştırabilmektedir. PCA algoritmasında ilk olarak eğitim kümesinin ortalama ortak değişinti matrisi Denklem 3'deki gibi hesaplanır.

$$C_V = \frac{1}{N} \sum_{i=1}^N F_i^T \times F_i \quad (25)$$

Bu denklemde N eğitim setindeki imge sayısı ve  $F_i$  i-inci imgenin öznitelik matrisidir. Eğitim kümesinin ortak değişinti matrisi  $C_V$  128×128 boyutunda bir matristir.  $C_V$ 'nin özdeğerleri  $\lambda_k$  (eigenvalue) ve özvektörleri  $\mathbf{u}_k$  (eigenvector) hesaplanmıştır. İki imgeyi karşılaştırmak için aşağıdaki adımlar izlenmiştir:

- i)  $F_m^T \times F_m$ 'nin kolonlarının ortalaması vektörü  $\boldsymbol{\gamma}$  ve  $F_n^T \times F_n$ 'nin kolonlarının ortalaması vektörü  $\boldsymbol{\zeta}$  olsun. Burada m referans imgenin indeksi ve n ise karşılaştırılacak imgenin indeksidir.
- ii)  $\boldsymbol{\gamma}$ 'yı  $\mathbf{u}_k$ 'nin ilk 20 vektörü (en büyük 20 özdeğere karşılık gelen 20 özvektör) ile çarparak Denklem 4'teki gibi referans imge 'hücre uzayı' üzerine izdüşürülür. Bu denklemde  $\mu_{\boldsymbol{\gamma}}$ ,  $\boldsymbol{\gamma}$ 'nın elemanlarının ortalamasıdır.

$$v_k = \mathbf{u}_k^T (\boldsymbol{\gamma} - \mu_{\boldsymbol{\gamma}}), \quad k = 1, 2, \dots, 20 \quad (26)$$

Benzer şekilde, karşılaştırılacak imge Denklem 5'teki gibi 'hücre uzayı' üzerine  $\boldsymbol{\zeta}$ 'yi  $\mathbf{u}_k$ 'nin ilk 20 vektörü (en büyük 20 özdeğere karşılık gelen 20 özvektör) ile çarparak izdüşürülür. Bu denklemde  $\mu_{\boldsymbol{\zeta}}$ ,  $\boldsymbol{\zeta}$ 'nin elemanlarının ortalamasıdır.

$$w_k = \mathbf{u}_k^T (\boldsymbol{\zeta} - \mu_{\boldsymbol{\zeta}}), \quad k = 1, 2, \dots, 20 \quad (27)$$

iii) İki imge arasındaki benzerlik ise şu şekilde tanımlanmıştır:

$$sim(I_1, I_2) = \cos(\mathbf{v}, \mathbf{w}) = \frac{\langle \mathbf{v} \cdot \mathbf{w} \rangle}{\|\mathbf{v}\|_2 \|\mathbf{w}\|_2} \quad (28)$$

Alternatif olarak bu benzerlik hesaplaması aşağıdaki şekilde de yapılabilir:

$$sim(I_1, I_2) = c4(\mathbf{v}, \mathbf{w}) = \frac{\langle \mathbf{v} \oplus \mathbf{w} \rangle}{2 \cdot \sum_{i=1}^{\{N\}} \max(|v_i|, |w_i|)} \quad (29)$$

Bu denklemdeki  $\oplus$  işlemi yukarıda tanımlanmıştır.

Yukarıda da belirtildiği üzere karaciğer imge kümesinin 3 sınıfa ait karaciğer doku imgelerini içerir. Derece I (17 hastadan alınan 119 imge), Derece II (17 hastadan alınan 151 imge) ve normal (9 hastadan alınan 184 imge) doku imgeleri. İmgeler  $300 \times 300$  boyutundadır. Sınama imgesinin sınıfını belirlemek için 3-en-yakın-komşu sınıflandırma algoritması kullanılmıştır. Eğitim kümesi oluşturulurken bir imgeyi dışarıda bırakmak yerine bir hastanın tüm imgelerini dışarıda bırakma yolu izlenmiştir. Bunun sebebi her hastaya ait birden fazla imgenin bulunmasıdır. Sadece imgeler için değil hastalar için de karar vermek için 'hasta sınıflandırma başarısı' kavramı tanımlanmıştır. Bu tanıma göre hastanın imgelerinin sınıflandırma sonuçları arasında çoğunluk oylaması yapılarak hasta hakkında karar verilir. Test kümesi bütün imgeleri içermesi için döndürülmüştür ve her döngüde bir imge içerir. Deneylede geri çağırma, duyarlılık ve genel doğruluk ölçüleri ile önerilen sistemin performansı ölçülmüştür.

$$Geri \text{ Çağırma} = \frac{\sum Gerçek \text{ Pozitifler}}{\sum Gerçek \text{ Pozitifler} + \sum Yanlış \text{ Negatifler}} \quad (30)$$

$$Duyarlılık = \frac{\sum Gerçek \text{ Pozitifler}}{\sum Gerçek \text{ Pozitifler} + \sum Yanlış \text{ Pozitifler}} \quad (31)$$

$$Genel \text{ Doğruluk} = \frac{\sum Gerçek \text{ Pozitifler} + \sum Gerçek \text{ Negatifler}}{\sum Tüm \text{ İmgeler}} \quad (32)$$

Deneylede PCA algoritması içerisinde iki farklı benzerlik ölçüsü kullanılmıştır. İlki geleneksel kosinüs benzerlik ölçüsü, ikincisi ise çarpmasız  $c4$  benzerlik ölçüsüdür.



**Tablo 12 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	70/119	41/119	8/119	58.8%
<i>Derece II</i>	38/151	102/151	11/151	67.5%
<i>Normal</i>	12/184	5/184	167/184	90.8%
<i>Duyarlılık</i>	58.3%	68.9%	89.8%	<i>Genel Doğruluk</i> 74.7%

**Tablo 13 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	11/17	5/17	1/17	64.7%
<i>Derece II</i>	3/17	12/17	2/17	70.6%
<i>Normal</i>	0/9	0/9	9/9	100%
<i>Duyarlılık</i>	78.6%	70.6%	75.0%	<i>Genel Doğruluk</i> 74.4%

Üç sınıf imge sınıflandırma probleminde %74.7 sınıflandırma başarısı elde edilmiştir. Aynı durumda hasta sınıflandırma başarısı ise %74.4 olmuştur. Normal imgelerin duyarlılık oranının %89.8'e kadar çıkabildiği görülmüştür. Hatalı sınıflandırılan imgelerin hastalara dağılımına bağlı olarak imge ve hasta sınıflandırma oranlarındaki farklar oluşabildiği gözlenmiştir.

**Tablo 14 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

	<i>Kanser</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Kanser</i>	252/270	18/270	93.3%
<i>Normal</i>	18/184	166/184	90.2%
<i>Duyarlılık</i>	93.3%	90.2%	<i>Genel Doğruluk</i> 92.1%

**Tablo 15 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

	<i>Heterojen</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Heterojen</i>	31/34	3/34	91.2%
<i>Normal</i>	0/9	9/9	100%
<i>Duyarlılık</i>	100%	75.0%	<i>Genel Doğruluk</i> 93.0%

İki sınıf 'Heterojen veya Normal' imge sınıflandırmasında %92.1, hasta sınıflandırmasında %93.0 sınıflandırma başarısı elde edilmiştir. Bu sonuçlar, sistemimizin doku imgesinde KKH olup olmadığını tespit etme konusunda oldukça başarılı olduğunu göstermiştir.

**Tablo 16 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	77/119	42/119	64.7%
<i>Derece II</i>	38/151	113/151	74.8%
<i>Duyarlılık</i>	67.0%	72.9%	<i>Genel Doğruluk</i> 70.4%

**Tablo 17 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	11/17	6/17	64.7%
<i>Derece II</i>	3/17	14/17	82.4%
<i>Duyarlılık</i>	78.6%	70.0%	<i>Genel Doğruluk</i> 73.5%

İki sınıf 'Derece I – Derece II' imge sınıflandırmasında %70.4, hasta sınıflandırmasında ise %73.5 (Tablo 17) sınıflandırma başarısı elde edilmiştir. Bu sonuçlar kanser kök hücreli imgelerin de kendi içlerinde ayırt edilebildiğini gösterdiği için önemlidir.

**Tablo 18 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	61/119	42/119	16/119	51.2%
<i>Derece II</i>	35/151	109/151	7/151	72.2%
<i>Normal</i>	4/184	5/184	175/184	95.1%
<i>Duyarlılık</i>	61.0%	69.9%	88.4%	<i>Genel Doğruluk 76.0%</i>

**Tablo 19 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	10/17	6/17	1/17	58.8%
<i>Derece II</i>	4/17	12/17	1/17	70.6%
<i>Normal</i>	0/9	0/9	9/9	100%
<i>Duyarlılık</i>	71.4%	66.7%	81.8%	<i>Genel Doğruluk 72.1%</i>

Üç sınıf sınıflandırma probleminde %76.0 imge sınıflandırma başarısı elde edilmiştir. Aynı durumda hasta sınıflandırma başarısı ise %72.2 olmuştur. Çarpmasız c4 benzerlik ölçüsü ile elde edilen imge sınıflandırma başarısının kosinüs benzerlik ölçüsü ile elde edilen sınıflandırma başarısından daha iyi olduğu gözlenmiştir. Üç sınıf hasta sınıflandırmasında ise c4 benzerlik ölçüsü ile elde edilen sonuçlar kosinüs benzerlik ölçüsü ile elde edilen sonuçlardan daha iyi değildir, fakat sonuçlar birbirine oldukça yakındır.

**Tablo 20 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

	<i>Heterojen</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Heterojen</i>	248/270	22/270	91.9%
<i>Normal</i>	10/184	174/184	94.6%
<i>Duyarlılık</i>	96.1%	88.8%	<i>Genel Doğruluk 93.0%</i>

**Tablo 21 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

	<i>Heterojen</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Heterojen</i>	30/34	4/34	88.2%
<i>Normal</i>	0/9	9/9	100%
<i>Duyarlılık</i>	100%	69.2%	<i>Genel Doğruluk</i> 90.7%

İki sınıf 'Heterojen veya Normal' sınıflandırma probleminde %93.0 imge sınıflandırma başarısı elde edilmiştir. Aynı durumda hasta sınıflandırma başarısı ise %90.7 olmuştur.

**Tablo 22 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	81/119	38/119	68.1%
<i>Derece II</i>	33/151	118/151	78.1%
<i>Duyarlılık</i>	71.1%	75.6%	<i>Genel Doğruluk</i> 73.7%

**Tablo 23 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	12/17	5/17	70.6%
<i>Derece II</i>	3/17	14/17	82.4%

#### **4.2 Akciğer İmge Seti Üzerinde Elde Edilen Deneysel Sonuçlar**

Akciğer imge kümesinin 3 sınıfa ait akciğer doku imgelerini içerir; Derece I (8 hastadan alınan 120 imge), Derece II (8 hastadan alınan 120 imge) ve normal (3 hastadan alınan 120 imge) doku imgeleri (Şekil 11). İmgeler 150×150 boyutundadır.

Bir imgenin derecesini belirlemek için üç en yakın komşu sınıflandırma algoritması kullanılmıştır. Eğitim kümesi oluşturulurken bir imgeyi dışarıda bırakmak yerine bir hastanın tüm imgelerini dışarıda bırakma yolu izlenmiştir. Bunun sebebi her hastaya ait birden fazla imgenin bulunmasıdır. Sadece imgeler için değil hastalar için de karar vermek için 'hasta sınıflandırma başarısı' kavramı tanımlanmıştır. Bu tanıma göre hastanın imgelerinin sınıflandırma sonuçları arasında çoğunluk oylaması yapılarak hasta hakkında karar

verilir. Test kümesi bütün imgeleri içermesi için döndürülmüştür ve her döngüde bir imge içerir. Deneylede geri çağırma, duyarlılık ve genel doğruluk ölçüleri ile önerilen sistemin performansı ölçülmüştür.

Deneylede değiştirilmiş PCA algoritması içerisinde iki farklı benzerlik ölçüsü kullanılmıştır. İlki geleneksel kosinüs benzerlik ölçüsü, ikincisi ise çarpmasız  $c_4$  benzerlik ölçüsüdür.

**Tablo 24 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	77/120	36/120	7/120	64.2%
<i>Derece II</i>	38/120	75/120	7/120	62.5%
<i>Normal</i>	9/120	5/120	106/120	88.3%
<i>Duyarlılık</i>	62.1%	64.7%	88.3%	<i>Genel Doğruluk</i> 71.7%

**Tablo 25 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	6/8	2/8	0/8	75.0%
<i>Derece II</i>	2/8	6/8	0/8	75.0%
<i>Normal</i>	0/3	0/3	3/3	100%
<i>Duyarlılık</i>	75.0%	75.0%	100%	<i>Genel Doğruluk</i> 79.0%

Üç sınıf imge sınıflandırma probleminde %71.7 sınıflandırma başarısı elde edilmiştir (Tablo 31). Aynı durumda hasta sınıflandırma başarısı ise %79.0 olmuştur (Tablo 32). Normal hastaların geri çağırma ve duyarlılık oranlarının %100 olduğu göz önüne alındığında, sistemin KKH içermeyen dokulara (normal dokular) koyduğu tanıların oldukça güvenilir olduğu söylenebilir.

**Tablo 26 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

	<i>Heterojen</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Heterojen</i>	227/240	13/240	94.6%
<i>Normal</i>	15/120	105/120	87.5%
<i>Duyarlılık</i>	93.8%	89.0%	<i>Genel Doğruluk</i> 92.2%

**Tablo 27 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

	<i>Heterojen</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Heterojen</i>	16/16	0/16	100%
<i>Normal</i>	0/3	3/3	100%
<i>Duyarlılık</i>	100%	100%	<i>Genel Doğruluk</i> 100%

İki sınıf ‘Heterojen veya Normal’ imge sınıflandırmasında %92.2 sınıflandırma başarısı elde edilmiştir. Hasta sınıflandırılmasında elde edilen %100 genel doğruluk oranı oldukça dikkat çekicidir. Bu sonucun ortaya çıkmasında hasta sınıflandırmada kullanılan hastanın imgeleri arasındaki çoğunluk oylamasının önemi çok büyüktür. Bu sonuç göstermektedir ki bir hastanın bir yerine birden fazla doku imgesi kullanıldığında sistemimiz daha başarılı sonuçlar elde etmektedir. Bu sebeple bir hastaya ait sadece bir imge olduğu durumlarda bile bu imgenin birden fazla parçaya ayrılarak sisteme girdi olarak verilmesi daha güvenilir sonuçlar üretmektedir ve sistemin bu şekilde kullanılması tavsiye edilmektedir.

**Tablo 28 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	77/120	43/120	64.2%
<i>Derece II</i>	39/120	81/120	67.5%
<i>Duyarlılık</i>	66.4%	65.3%	<i>Genel Doğruluk</i> 65.8%

**Tablo 29 Kosinüs Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	6/8	2/8	75.0%
<i>Derece II</i>	2/8	6/8	75.0%
<i>Duyarlılık</i>	75.0%	75.0%	<i>Genel Doğruluk</i> 75.0%

İki sınıf ‘Derece I – Derece II’ imge sınıflandırmasında %65.8, hasta sınıflandırmasında ise %75.0 sınıflandırma başarısı elde edilmiştir. Bu sonuçlar kanser kök hücreli

doku imgelerinin de kendi içlerinde ayırt edilebildiğini gösterdiği için önemlidir. Ayrıca hasta sınıflandırma başarı oranının imge sınıflandırmadan yaklaşık %10 daha iyi olduğu gözlenmiştir. Bu da bir önceki durumda olduğu gibi hastaya ait birden fazla imge kullanımını teşvik eden bir sonuçtur.

**Tablo 30 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	69/120	42/120	9/120	57.5%
<i>Derece II</i>	40/120	75/120	5/120	62.5%
<i>Normal</i>	12/120	9/120	99/120	82.5%
<i>Duyarlılık</i>	57.0%	59.5%	87.6%	<i>Genel Doğruluk</i> 67.5%

**Tablo 31 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 3 Sınıf 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	5/8	3/8	0/8	62.5%
<i>Derece II</i>	2/8	6/8	0/8	75.0%
<i>Normal</i>	0/3	0/3	3/3	100%
<i>Duyarlılık</i>	71.4%	66.7%	100%	<i>Genel Doğruluk</i> 79.0%

Üç sınıf imge sınıflandırma probleminde %67.5 sınıflandırma başarıları elde edilmiştir. Aynı durumda hasta sınıflandırma başarıları ise %79.0 olmuştur. Normal hastaların geri çağırma ve duyarlılık oranlarının %100 olduğu göz önüne alındığında, sistemin KKH içermeyen dokulara (normal dokular) koyduğu tanıların oldukça güvenilir olduğu söylenebilir. Bu deneylerde c4 benzerlik ölçüsü kosinüs benzerlik ölçüsünün ürettiği sonuçlara oldukça yakın sonuçlar üretmektedir.

**Tablo 32 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

	<i>Heterojen</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Heterojen</i>	226/240	14/240	94.2%
<i>Normal</i>	22/120	98/120	81.7%
<i>Duyarlılık</i>	91.1%	87.5%	<i>Genel Doğruluk</i> 90.0%

**Tablo 33 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Heterojen veya Normal) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

	<i>Heterojen</i>	<i>Normal</i>	<i>Geri Çağırma</i>
<i>Heterojen</i>	16/16	0/16	100%
<i>Normal</i>	0/3	3/3	100%
<i>Duyarlılık</i>	100%	100%	<i>Genel Doğruluk</i> 100%

İki sınıf 'Heterojen veya Normal' imge sınıflandırmasında %90.0, hasta sınıflandırmasında %100 sınıflandırma başarısı elde edilmiştir. Yukarda da belirtildiği üzere hasta sınıflandırılmasında elde edilen %100 genel doğruluk oranı oldukça dikkat çekicidir. Bu durumda c4 benzerlik ölçüsü ile elde edilen imge sınıflandırma başarısı kosinüs benzerlik ölçüsü ile elde edilen başarıya oldukça yakındır. Hasta sınıflandırmada ise iki benzerlik ölçüsü aynı sonuçları üretmiştir.

**Tablo 34 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu İmge Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	73/120	47/120	60.8%
<i>Derece II</i>	35/120	85/120	70.8%
<i>Duyarlılık</i>	67.6%	64.4%	<i>Genel Doğruluk</i> 65.8%



**Tablo 35 c4 Benzerliği Kullanan Değiştirilmiş PCA Algoritması ile 2 Sınıf (Derece I- Derece II) 3 En Yakın Komşu Hasta Sınıflandırmasının Karıştırma Matrisi**

<i>Heterojenite</i>	<i>Derece I</i>	<i>Derece II</i>	<i>Geri Çağırma</i>
<i>Derece I</i>	5/8	3/8	62.5%
<i>Derece II</i>	1/8	7/8	87.5%
<i>Duyarlılık</i>	83.3%	70.0%	<i>Genel Doğruluk</i> 75.0%

İki sınıf “Derece I – Derece II” imge sınıflandırmasında %65.8, hasta sınıflandırmasında ise %75.0 sınıflandırma başarısı elde edilmiştir. Bu sonuçlar kanser kök hücreli doku imgelerinin c4 benzerlik ölçüsü ile de kendi içlerinde ayırt edilebildiğini gösterdiği için önemlidir. Ayrıca kosinüs ile elde edilen sonuçlara paralel olarak bu sonuçlarda da hasta sınıflandırma başarı oranının imge sınıflandırmadan yaklaşık %10 daha iyi olduğu gözlenmiştir. Bu da deneysel sonuçların büyük çoğunluğunda olduğu gibi hastaya ait birden fazla imge kullanımını teşvik eden bir sonuçtur.

Projemizde H&E boyanmış akciğer ve karaciğer doku imgelerini sınıflandırabilen bir algoritma geliştirilmiştir. H&E boyanmış doku imgelerinde KKH'ler çıplak gözle görülemediği için önerilen değiştirilmiş PCA algoritması bir örüntü sınıflandırma algoritmasıdır. Üç sınıfa ait 454 imge içeren karaciğer imge setinde çarpmasız benzerlik ölçüsü kullanıldığında %76.0 sınıflandırma başarısı elde edilmiştir. İki sınıf ‘Heterojen veya Normal’ sınıflandırma probleminde ise başarı oranı %93.0'e kadar çıkmaktadır. İki sınıf kanserli dokuların kanser-kök-hücre heterojen imgelerin “Derece I – Derece II” sınıflandırmasında ise %76.5 başarı oranı elde edilmiştir. Bu sonuçlar üç sınıf dokunun başarılı bir şekilde ayrıştırılabildiğini göstermektedir. Akciğer imge setinde de benzer sınıflandırma performansları gözlenmiştir. Ayrıca akciğer imge setinde elde edilen %100 ‘Heterojen veya Normal’ hasta sınıflandırma sonucu hem hastaya ait birden fazla doku imgesi kullanmanın önemini hem de önerilen algoritmanın başarısını göstermektedir.

### **İmmünohistokimya İmgeleri Üzerinde Yapılan Deneysel Çalışmalar**

Bu imge grubu üzerinde yapılan deneysel çalışmada da Matthews Korelasyon Katsayısı (MKK) <sup>16</sup>, hassasiyet ya da keskinlik (precision), tanıma oranı, F1 skoru (Yang ve Liu, 1999) ve özgüllük değerleri hesaplanmıştır.

Özgüllük, Tanıma, Keskinlik ve F1 tanımları da aşağıdaki gibidir:

$$\text{Özgüllük} = \frac{TN}{TN + FP} \quad 33$$

$$\text{Tanıma} = \frac{TP}{TP + FN} \quad 34$$

$$\text{Keskinlik} = \frac{TP}{TP + FP} \quad 35$$

$$F1 = 2 * \frac{\text{Keskinlik} * \text{Tanıma}}{(\text{Keskinlik} + \text{Tanıma})} \quad 36$$

Matthew Korelasyon Katsayısı aşağıdaki gibi hesaplanmıştır:

$$MKK = \frac{(TP)(TN) - (FP)(FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad 37$$

bu denklemdaki TP, TN, FP ve FN elemanları sırasıyla gerçek pozitif, gerçek negatif, yanlış kabul ve ikinci tür hatayı temsil etmektedir. Aşağıdaki Tabloda Şekil 19'daki imgeyi içeren slayt için verilmiştir. Sınıflandırıcılar elimizdeki diğer resimler kullanılarak eğitilmiştir.

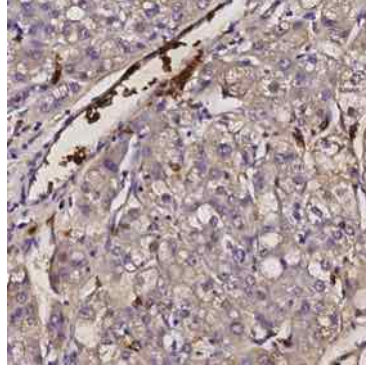
**Tablo 36 Şekil 10'daki imgenin sonuçları**

	MKK	F1	Özgüllük	Tanıma	Keskinlik
Ortak Fark DVM	0.231	0.1683	0.88055	0.71433	0.095389
Kovaryans DVM	0.13407	0.075518	0.6241	0.87576	0.03946
Ortak Fark YSA	0.13621	0.15176	0.97873	0.18115	0.13058
RGB Ort. YSA	0.46077	0.46924	0.99179	0.44934	0.49098
Kovaryans YSA	0.21523	0.21259	0.99384	0.16047	0.31488
RGB Ort. DVM	0.47704	0.43592	0.96682	0.80314	0.29914
YCbCr DVM	0.50971	0.46307	0.96737	0.85893	0.31699
YCbCr YSA	0.61274	0.619	0.992	0.65165	0.58947
İlk Sonuç	0.59046	0.59644	0.99091	0.64396	0.55545
Eğitilmiş Sonuç	0.68351	0.66537	0.9856	0.90574	0.52583

Bu resimde KKH tanıma oranı kullanıcı tarafından doğrulama yapılmasından sonra %90'nin üstüne çıkmaktadır. MKK ve F1 oranlarında da iyileşme sağlanmaktadır. İlk aşamada en iyi tanıma oranı kovaryans matrisine dayalı destek vektör makinası ile elde edilmektedir ancak bu sınıflandırıcının hassasiyeti, F1 ve MKK'si de çok düşüktür. Yani KKH'lerin olmadığı yerleri de KKH diye işaretlemektedir. Betimleyici karıştırma metodu ile eğitime sonrasında hem tanıma oranı artmakta hem de MKK, F1 ve hassasiyet artmaktadır. Bir sonraki

Tablo'da başka bir resim için sonuçlar vardır. Burada da ikincil eğitime sonrasında tanıma oranında yükselme sağlanmıştır. MKK, F1 ve keskinlik oranları da artırılmıştır.

Aşağıdaki imgeden de görüleceği üzere özellikle boşluk kenarlarında kahverengi boyalı bölgeler oluşmaktadır. Yazılım bu bölgeleri de işaretlemektedir. Bu bölgeler hücre çekirdeği olmadığından kolaylıkla gözle elenebilirler.



**Şekil 10 CD13 kanser kök hücre işaretleyicisiyle boyanmış karaciğer kanser dokusu**

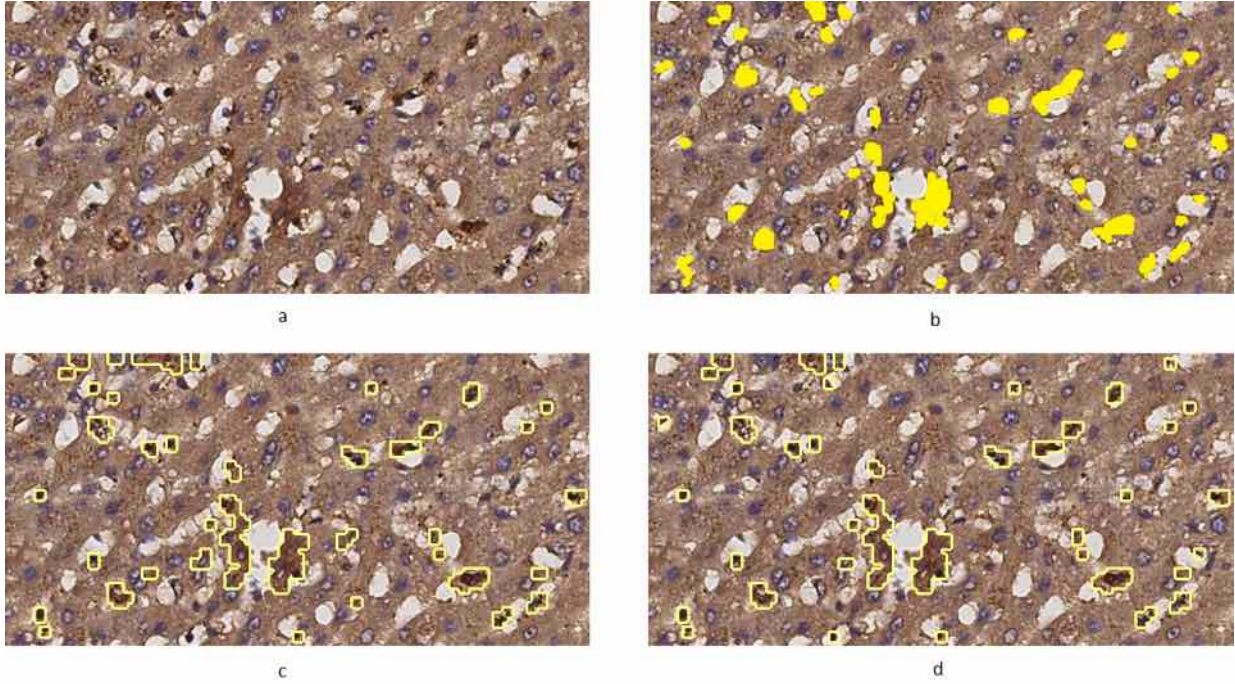
**Tablo 37 Şekil 11'deki imgenin sonuçları**

	MKK	F1	Özgüllük	Tanıma	Keskinlik
Ortak Fark DVM	0.19251	0.17779	0.95247	0.40155	0.11417
Kovaryans DVM	0.13171	0.093444	0.84016	0.56258	0.050954
Ortak Fark YSA	0.15368	0.1603	0.99205	0.13256	0.20271
RGB Ort. YSA	0.31438	0.30405	0.99573	0.22943	0.45062
Kovaryans YSA	0.20632	0.21617	0.9904	0.19745	0.23882
RGB Ort. DVM	0.4927	0.49001	0.98584	0.62581	0.40264
YCbCr DVM	0.45861	0.46166	0.98714	0.55315	0.39613
YCbCr YSA	0.52353	0.52369	0.99566	0.45572	0.61548
İlk Sonuç	0.46712	0.46091	0.99613	0.37549	0.59665
Eğitilmiş Sonuç	0.64666	0.65194	0.99426	0.66574	0.6387

Diğer CD13 görüntülerinde de benzer sonuçlar elde edilmektedir.

CANSTEM yazılımının nasıl çalıştığını gösteren bir video hazırlanmıştır. Bu videoda eğitilmiş katsayılar kullanılarak bir CD13 imgesindeki kanser kök hücrelerinin bulunması gösterilmektedir. Bilindiği üzere projenin amacı immünohistokimya (IHK) yöntemiyle boyanmış imgelerde kanser kök hücre tespit etmek değildir. Bu yazılım bizim açımızdan yer gerçekliği oluşturmak amacıyla kullanılmıştır. Ayrıca patoloğlar tarafından da KKH saymak amacıyla kullanılabilir. CANSTEM yazılımını ImageJ ve Fiji hazır programlarının eklentiler ile de karşılaştırdık. Geliştirilen yazılımın sonuçları, ImageJ ve Fiji programlarının eklentilerine

göre %20 daha iyi sonuç vermektedir. Aşağıdaki tablolarda detaylı karşılaştırma sonuçları verilmiştir.



**Şekil 11 (a) CD13 ile boyanmış bir imge, (b) aynı imgenin kesin referans, (c) betimleyici karıştırma metodunun ilk sonucu, (d) kullanıcı geri dönüşü alındıktan sonraki sonuç.**

CANSTEM yazılımının çıktısı Şekil 11'de CD13 ile boyanmış karaciğer dokusu için gösterilmiştir. Kesin referans ile karşılaştırıldığında yanlış işaretlenen yerler ile işaretlenmesi gerekip işaretlenmemiş yerler görülmektedir. Şekil 11(d)'de kullanıcı geri dönüşü alındıktan sonra bulunan kanser kök hücreleri gösterilmiştir. Görüleceği üzere hatalar azalırken kesin referansa yakın bir sonuç elde edilmektedir. Geliştirilen algoritma, pataloglar tarafından sıklıkla kullanılan ImageJ programının Renksel Bölütleme (Color Segmentation) eklentisi ile karşılaştırılmıştır (<http://bigwww.epfl.ch/sage/soft/colorsegmentation/>). Renksel Bölütleme eklentisinde, kullanıcı bölütlemek istediği bölgeleri kendisi seçer. Eklenti 10 adet farklı bölge seçilmesine imkan tanımaktadır. Seçilen bölgeler K-Ortalama ve Gizli Markov Modeli kümeleme algoritmalarını kullanmaktadır. Deneylerimizde zamansal kısıtları dahil ettiği için Gizli Markov Modeli kullanılmıştır. Gizli Markov Modelinde ortak ve ayrı renk düzenleri için ayrı deneyler yapılmıştır.

Betimleyici karıştırma metodu (BKM), ImageJ renksel bölütleme eklentisine göre daha iyi sonuç vermektedir (Tablo 38). Ancak karşılaştırılan bu eklenti ile bir sınıflandırıcı eğitmek mümkün değildir. Fiji programının Eğitilebilir Weka Bölütleme (EWB) eklentisi (<http://imagej.net/Trainable Weka Segmentation>) ise bu amaç için kullanılmaktadır. EWB

eklentisinin amacı Weka programının makine öğrenmesi gücünü Fiji ile birleştirmektir. EWB'de kullanıcı tarafından belirtilen örnek bölgelerden öznitelikler çıkartılır. Bu öznitelikler ile sınıflandırıcılar eğitilir. Deneylerimizde BKM algoritması ile EWB eklentisi karşılaştırılmıştır. EWB eklentisinde çıkartılan öznitelikler, BKM algoritmasında çıkartılan özniteliklere yakın olmaları nedeniyle, ortalama renk ve türev değerleri olacak şekilde seçilmiştir. Eğitilen yöntem ise Hızlı-Rastlantısal Orman algoritmasıdır. Bu algoritma EWB eklentisinin öntanımlı eğitim yöntemi olup, 200 ağaç üstünden her dal için 2 rastlantısal öznitelik almaktadır.

Tablo 39'da , BKM ve EWB eklentisinin MKK ve F1 skorları verilmiştir. Bazı imgelerde EWB de beklenenden daha düşük sonuçlar elde etmiştir. Bunun nedeni bu eklentide eğitilen makine öğrenmesinin tek imge üzerinden örneklenmesidir. Bu yüzden aynı boya ile boyanmış farklı bir imge bölütlenmeye çalışıldığında eğitilen makine bu değişime cevap verememektedir. Ancak, geliştirdiğimiz Betimleyici karıştırma metodu bu tür değişimlere karşı dayanıklıdır ve daha iyi bölütleme sonuçları vermektedir

**Tablo 38 CANSTEM Yazılımında kullanılan algoritma ile ImageJ renksel bölütleme eklentisinin karşılaştırılması**

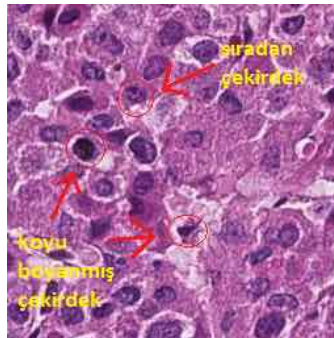
İmge #	BKM			GMM Ayrı Renk Düz.			GMM Ortak Renk Düz.		
	MKK	F1	Özg.	MKK	F1	Özg.	MKK	F1	Özg.
CD13-1	0.61	0.64	0.95	0.35	0.33	0.99	0.35	0.33	0.99
CD13-2	0.61	0.61	0.96	0.40	0.47	0.97	0.40	0.47	0.97
CD13-3	0.68	0.68	0.96	0.45	0.45	0.97	0.45	0.45	0.98
CD13-4	0.63	0.62	0.96	0.43	0.45	0.97	0.48	0.47	0.99
CD13-5	0.60	0.61	0.91	0.46	0.46	0.98	0.48	0.49	0.99
CD13-6	0.63	0.63	0.95	0.46	0.48	0.97	0.47	0.48	0.98
CD13-7	0.58	0.56	0.95	0.48	0.49	0.97	0.47	0.48	0.98
CD13-8	0.53	0.52	0.98	0.97	0.38	0.99	0.40	0.39	0.99
CD13-9	0.58	0.58	0.99	0.44	0.45	0.99	0.42	0.41	0.98
CD13-10	0.65	0.65	0.99	0.43	0.43	0.98	0.42	0.43	0.98
CD13-11	0.63	0.63	0.99	0.55	0.55	0.99	0.44	0.44	0.99
CD13-12	0.70	0.69	0.99	0.48	0.49	0.99	0.45	0.44	0.99
CD13-13	0.63	0.61	0.98	0.46	0.45	0.99	0.49	0.46	0.99
CD13-14	0.59	0.59	0.99	0.35	0.30	0.97	0.23	0.17	0.96
CD13-15	0.60	0.60	0.99	0.27	0.20	0.99	0.41	0.39	0.98
CD13-16	0.67	0.65	0.97	0.61	0.62	0.98	0.56	0.56	0.97
CD13-17	0.68	0.67	0.98	0.42	0.42	0.98	0.28	0.24	0.94
CD13-18	0.63	0.61	0.99	0.43	0.44	0.99	0.39	0.39	0.99

**Tablo 39 CANSTEM’de kullanılan BKM ile EWB metotlarının karşılaştırılması**

İmge #	Betimleyici Karıştırma Metodu		EWB Hızlı-Rastlantısal Orman	
	MKK	F1	MKK	F1
CD13-1	0.61	0.64	0.36	0.36
CD13-2	0.61	0.61	0.42	0.48
CD13-3	0.68	0.68	0.46	0.48
CD13-4	0.63	0.62	0.46	0.48
CD13-5	0.60	0.61	0.48	0.49
CD13-6	0.63	0.63	0.46	0.47
CD13-7	0.58	0.56	0.47	0.49
CD13-8	0.53	0.52	0.38	0.38
CD13-9	0.58	0.58	0.25	0.19
CD13-10	0.65	0.65	0.22	0.18
CD13-11	0.63	0.61	0.55	0.56
CD13-12	0.70	0.69	0.56	0.56
CD13-13	0.63	0.61	0.60	0.61
CD13-14	0.59	0.59	0.54	0.54
CD13-15	0.60	0.60	0.63	0.64
CD13-16	0.67	0.65	0.63	0.42
CD13-17	0.68	0.67	0.60	0.61
CD13-18	0.63	0.61	0.17	0.12

12.Şekil’de görüldüğü üzere H&E boyanmış slaytlarda bazı çekirdekler diğer çekirdeklere göre daha koyu boyanmakta ve yanlarında boşluklar oluşmaktadır. Bu nedenle de imgelerin örüntülerinde farklılıklar oluşmaktadır. Geliştirdiğimiz algoritma ve yazılım da bu farklılıklardan faydalanarak imgeleri sınıflandırabilmektedir.

H&E ile boyanmış imgelerdeki koyu boyanmış çekirdeklerle CD13 ile boyanmış imgelerdeki kahverengi boyalı (büyük ihtimalle kanser kök hücresi olan) hücreler arasında bir ilişki gözlemledik.



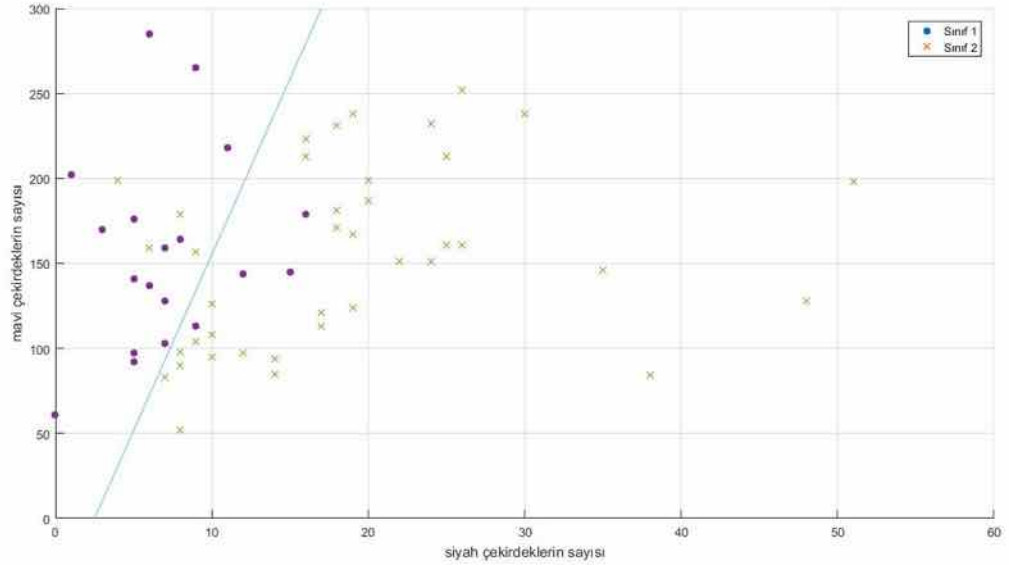
**Şekil 12: H&E ile boyanmış karaciğer kanseri imgesi. Koyu boyanmış çekirdekler büyük ihtimalle kanser kök hücreleridir.**

Kanserli CD13 slaytlarını kahverengiye boyanmış çekirdeklerin maviye boyanmış çekirdek oranına göre iki kategoriye ayırdık. Bu oranın yüzdesine KBO diyelim.

$KBO = 100 \times (\text{Kahverengiye boyanmış çekirdeklerin sayısı}) / (\text{Maviye boyanmış çekirdeklerin sayısı})$

KBO oranı ile kök hücre oranının ilişkili olduğu literatürde belirtilmiştir (Yamashita vd., 2013).

3. Şekilde (x) ile işaretlenmiş görüntüler CD13 ile boyanmış slaytlardaki KBO oranının %3'ten büyük olduğu slaytları göstermektedir. Aynı şekilde (•) ile işaretlenmiş olanlar ise oranının %3'ten küçük olduğu slaytları göstermektedir. 13. Şekilden de görülebileceği üzere H&E resimlerinde siyah ya da koyu boyanmış çekirdek oranı ile KBO arasında ciddi bir ilişki vardır. Eğer H&E boyalı imgelerde koyu boyanmış çekirdeklerin diğer çekirdeklere oranı ile bir sınıflandırma yapılırsa CD13 sınıflandırmasına göre doğruluk oranı %86.2 çıkmaktadır. Başka hastalarda da benzer oranlara ulaşılmıştır.



**Şekil 13: H&E ile boyanmış imgelerdeki koyu ve mavi çekirdek sayıları. Yatay ekseninde koyu çekirdek sayısı, dikey ekseninde ise sıradan bir şekilde maviye boyanmış çekirdeklerin sayısı vardır. Sınıf 1 CD13 ile boyanmış slaytlardaki KBO oranının %3'ten az olan imgeleri göstermektedir. Sınıf 2 ise %3'ten fazla olan imgeleri göstermektedir. Bu grafik oluşturulurken H&E slaytlarının 4 değişik bölgesinden 256X256'lık parçalar alınmıştır.**

Yukardaki şekilden de görüleceği üzere KBO oranı yüksek hastalarda (yani kök hücre oranı yüksek imgelerde) koyu ya da siyah boyanmış çekirdek oranı fazla olmaktadır. Genel olarak

da kanser kök hücresi sayısı fazla ise H&E imgelerdeki koyu boyalı ya da siyah boyalı çekirdek sayısı fazla olmaktadır. Bu gözlemden yola çıkarak yaptığımız çalıştayda patoloğlara eğer H&E imgelerde normalden fazla koyu renkli çekirdek gözlemlemişlerse karaciğer kanserli hastalarda CD13 boyaması da yaptırılmalarını tavsiye ettik.

### 4.3 Görüntü Analizi ve Dijital Patoloji Sempozyumu

Proje çıktısı CANSTEM yazılımının patoloğlara ve dijital patoloji alanında çalışan araştırmacılara tanıtılması projenin önemli bir hedefiydi. Bu amaçla 29 Kasım 2016 tarihinde, Orta Doğu Teknik Üniversitesi, Enformatik Enstitüsünde, Türkiye Sağlık Enstitüleri (TÜSEB) Başkanı Prof. Dr. Fahrettin Keleştemur'un açılış konuşmasını gerçekleştirdiği ve uluslararası konuşmacıların katılımı ile Dijital Patoloji Sempozyumu gerçekleştirilmiştir. Sempozyum <http://users.metu.edu.tr/rengul/sempozyum.html> web adresinden duyurulmuştur. Sempozyuma 80 dinleyici katılmıştır (Ek2). Sempozyum katılımcılarına CANSTEM yazılımı hafıza çubuğu ile iletilmiştir. Sempozyum başlıkları aşağıda listelenmiştir.

“Hoşgeldiniz Konuşması”

Prof Dr Gönül Turhan Sayan, ODTU, Vice President for Research

“Açılış konuşması”

Prof Dr. Fahrettin Keleştemur, Türkiye Sağlık Enstitüleri Başkanı

“Digital pathology and its applications”

Ayşegül Üner, Hacettepe Üniv

“A Perspective on Digital Pathology and Image Analysis”

Christian Munzenmayer, Fraunhofer IIS

“Mobile phone based microscopes as digital imaging, sensing, and diagnostic tools”

Hatice Ceylan Koydemir, UCLA

“Aperio Digital Pathology”

Selim Effat, LEICA

“Tumor Heterogeneity and Image analysis”

A. Enis Çetin, Bilkent Üniv.

“Deep learning in digital pathology”

Uğur Halıcı, ODTÜ



## 5. TARTIŞMA VE SONUÇ

Mikroskopik görüntüler modern tıpta kanser tanısının önemli bir parçasıdır. Fakat, dokulara mikroskop altında tanı koymak patologlar için oldukça zaman alıcı ve yorucu bir işittir. Sağlık Bakanlığı kayıtlarına göre Ülkemizde 1200 patolog çeşitli kurumlarda görev almaktadır. Bu rakam genel nüfusa oranla oldukça düşük bir orandadır. Bu durum Dünya da farklı değildir. Benzer nüfusa sahip Almanya'da yaklaşık 1100 patolog bulunmaktadır. Bu bağlamda hedeflerini başarı ile tamamladığımız bu proje de patologlar için kanser işlemlerini hızlandırma amacıyla H&E boyanmış doku imgelerini sınıflandıran ve derecelendiren bilgisayar yardımlı bir tanı sistemi geliştirdik. Bu sistemimizde ilk adım doku imgelerinin etiketlenmesi gerçekleştirilmiştir ve yer belirteci imgeler oluşturulmuştur. Proje kapsamında imgelerin etiketlenmesi oldukça zaman alan ve önem gerektiren bir işlemdir. Kök hücre belirteçleri ile boyanmış imgeler yer gerçekliği etiketlerinin oluşturulması için kullanılmıştır, çünkü H&E boyanmış dokularda kanser kök hücreleri (KKH) çıplak gözle görülememektedir. Sistemimizdeki sonraki adım öznitelik çıkarımıdır. KKH'ler H&E boyanmış dokularda çıplak gözle görülemediği için örüntü ayırt edici öznitelikler çıkarılması gerekmektedir. Araştırma gurubunda yer alan biyologlar ve patoloğumuz ile beraber kanser kök hücre IHC (CD13 veya CD166) ve H&E ile boyanmış seri kesitlerde ortak özellikler belirlenmiş ve bu amaçla dokuz farklı renk uzayından 20 öznitelik seçilmiştir.

Bu öznitelikler, bu projede kovaryans matrisine dayalı sınıflandırıcılara ve Ana Bileşenler Analizi (PCA) algoritmasına girdi olarak verilmiştir. Ana bileşenler algoritmasının üç sınıf H&E boyanmış karaciğer dokularını sınıflandırma probleminde 76.0% imge sınıflandırma başarısı sağladığı deneysel sonuçlarla gösterilmiştir. Ana Bileşenler Analizi kanserli durum belirleme probleminde ise %90'ın üstünde başarı sağlamaktadır. Kovaryans matrisine dayalı sınıflandırıcılar ise üç sınıf ayırıştırma probleminde %90'a yakın başarı sağlamaktadırlar.

Ayrıca H&E imgelerde koyu renkli çekirdek sayısı ile CD13 boyanmış imgelerdeki kahverengi boyalı çekirdekler (muhtemel kök hücreleri) arasında bir ilişki gözlenmiştir. Kanımızca eğer H&E imgelerde koyu renkli çekirdek sayısı fazlaysa CD13 boyama da yaptırılmalı ve KKH oranı daha hassas bir şekilde belirlenmelidir. Bu ilişkiyi doğrulamak için aynı slaytın hem CD13 hem de H&E ile boyanması gereklidir. Bu deneysel çalışmada ileride yapılacaktır.

Bu proje kapsamında en sık görülen epitelyal kanserler hedeflenmiş ve hematolojik malignansiler kapsanmamıştır. Gelecekte benzer uygulamalar bu alanda da gerçekleştirilebilir. Ayrıca bu uygulama karaciğer, akciğer ve meme kanseri dokularının yanı sıra diğer prostat kolon tiroit gibi diğer sıkça görülen kanserler için de denenebilir ancak öncelikle etiketlenmiş imge setlerinde test edilmesi gerekir.

Sonuç olarak proje kapsamında geliştirilen CANSTEM öngörü aracı web üzerinden indirilebilir, (<http://users.metu.edu.tr/rengul/canstem.html>), bir yazılım paketi olarak kullanıma hazırdır. Her ne kadar proje sonuçlarının duyurulmasına yönelik proje son 3 ayı içerisinde planladığımız çalıştayın Temmuz ve Ağustos 2016 aylarına denk gelmesi nedeniyle ve talep ettiğimiz proje uzatma süresini alamadığımız için proje danışmanın önerisini gerçekleştiremedik. Ancak rutin patolojik preparatlarda kanser kök hücre oranını tanımaya yarayacağını düşündüğümüz klinik patolojik ve onkolog hekimlerin bilgisine 42. Ulusal Hematoloji Kongresinde (19-22 Ekim 2016) CANSTEM yazılımını ana hatlarıyla sunulmuş ve örnekler üzerinden nasıl çalıştığı katılımcılara gösterilmiştir. CANSTEM kanserli hastaların tedavisinde kişiye özgü yol haritası oluşturmalarını sağlayacak şekilde tasarlanmış olup ve tedavinin gecikmeden en etkin bir şekilde gerçekleştirilmesi gereken kanser hastalığında hem hesaplı hem de hızlı bir şekilde kullanılabilir. Onkologlar ve patolojiler sayısal patoloji uygulamalarını ve öngörü aralarını kullanarak ölümcül kanserli hastalara kişiye özgü kanser kök hücrelerini hedef alan tedavi yol haritası sağlayabilir. Çalıştayda hastanın H&E slaytlarında koyu çekirdek oranı fazlaysa kanser kök hücre oranının fazla olabileceği ve bu nedenle CD13 boyaması da yaptırılması gerektiği patolojilere tavsiye edilmiştir.

## 6. KAYNAKLAR LİSTESİ VE ATIFLAR

- 1 Correnti M, Raggi C. Stem-like plasticity and heterogeneity of circulating tumor cells: current status and prospect challenges in liver cancer. *Oncotarget* 2016; published online Oct 11. DOI:10.18632/oncotarget.12569.
- 2 Dalerba P, Cho RW, Clarke MF. Cancer stem cells: models and concepts. *Annu Rev Med* 2007; **58**: 267–84.
- 3 Zhou B-BS, Zhang H, Damelin M, Geles KG, Grindley JC, Dirks PB. Tumour-initiating cells: challenges and opportunities for anticancer drug discovery. *Nat Rev Drug Discov* 2009; **8**: 806–23.
- 4 Peinado H, Olmeda D, Cano A. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer* 2007; **7**: 415–28.
- 5 Oishi N, Wang XW. Novel therapeutic strategies for targeting liver cancer stem cells. *Int J Biol Sci* 2011; **7**: 517–35.
- 6 Fact Sheets by Cancer. [http://globocan.iarc.fr/Pages/fact\\_sheets\\_cancer.aspx](http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx).
- 7 O'Flaherty JD, Barr M, Fennell D, *et al.* The cancer stem-cell hypothesis: its emerging role in lung cancer biology and its relevance for future therapy. *J Thorac Oncol* 2012; **7**: 1880–90.
- 8 Iqbal J, Chong PY, Tan PH. Breast cancer stem cells: an update. *J Clin Pathol* 2013; **66**: 485–90.
- 9 Liu L-L, Fu D, Ma Y, Shen X-Z. The power and the promise of liver cancer stem cell markers. *Stem Cells Dev* 2011; **20**: 2023–30.
- 10 Tuzel O, Porikli F, Meer P. Region Covariance: A Fast Descriptor for Detection and Classification. Springer Berlin Heidelberg, 2006: 589–600.
- 11 Cosatto E, Miller M, Graf HP, Meyer JS. Grading nuclear pleomorphism on histological micrographs. In: 2008 19th International Conference on Pattern Recognition. IEEE, 2008: 1–4.
- 12 Suhre A, Alexander. Novel methods for microscopic image processing, analysis, classification and compression. 2013.
- 13 Tuna H, Onaran I, Çetin AE. Image Description Using a Multiplier-Less Operator. *IEEE Signal Process Lett* 2009; **16**. DOI:10.1109/LSP.2009.2024589.
- 14 Yang Y, Liu X. A re-examination of text categorization methods. In: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '99. New York, New York, USA: ACM Press, 1999: 42–9.
- 15 Turk M, Pentland A. Eigenfaces for recognition. *J Cogn Neurosci* 1991; **3**: 71–86.
- 16 Matthews BW. Comparison of the predicted and observed secondary structure of T4

phage lysozyme. *Biochim Biophys Acta* 1975; **405**: 442–51.

## 7. EK-1

### CANSTEM

#### Kanser Kök Hücresi Tespit Yazılımı

Kanserli dokuların mikroskop görüntülerinde kanser kök hücre oranının otomatik olarak belirlenmesi amacıyla tasarlanmıştır.

Kanser anormal hücrelerin kontrolsüz çoğalması ve yayılması olarak tanımlanan karmaşık bir hastalıktır. Bu nedenle kanser dokusu farklı özellikleri olan-ek popülasyon denilen-hücre gruplarını içerir. Bu gruplardan birisi olan kanser kök hücre popülasyonu kanserin kötü prognozundan sorumlu hücreler olarak tanımlanmaktadır. Hastalardaki hücre popülasyonunun birbirine göre oranı hastaların sağ-kalımında belirleyicidir. CANSTEM yazılımı TÜBİTAK (213E032) tarafından desteklenen bu proje kapsamında kanserin kemoterapötik tedavisinde izlenecek yol haritasında yönlendirici olması amacıyla geliştirilmiş ve patoloğların kullanımına sunulmuştur.

CANSTEM yazılımı Prof Dr. Enis Çetin, Prof Dr. Ayşegül Üner ve Doç. Dr. Rengul Atalay araştırma grupları tarafından geliştirilmiştir.

#### **Yükleme Yönergeleri:**

##### **Giriş:**

Bu kısım, "CSCDetection" yazılımının Windows ve OSX işletim sistemlerine indirilmesi ve yüklenmesine dair bilgiler içermektedir.

##### **Windows:**

*Eğer MATLAB R2016a ya da daha sonra ki versiyonlarınız var ise:*

- 1- "CSCDetection\_tr.exe"yi indirin.
- 2-"CSCDetection.exe"yi çalıştırabilirsiniz.

*Eğer MATLAB R2016a ya da daha sonra ki versiyonlarınız yok ise:*

- 1- MCR dosyasını indirin.
- 2- MCR'yi dosyayı çalıştırarak yükleyin.
- 3-Programı çalıştırın.

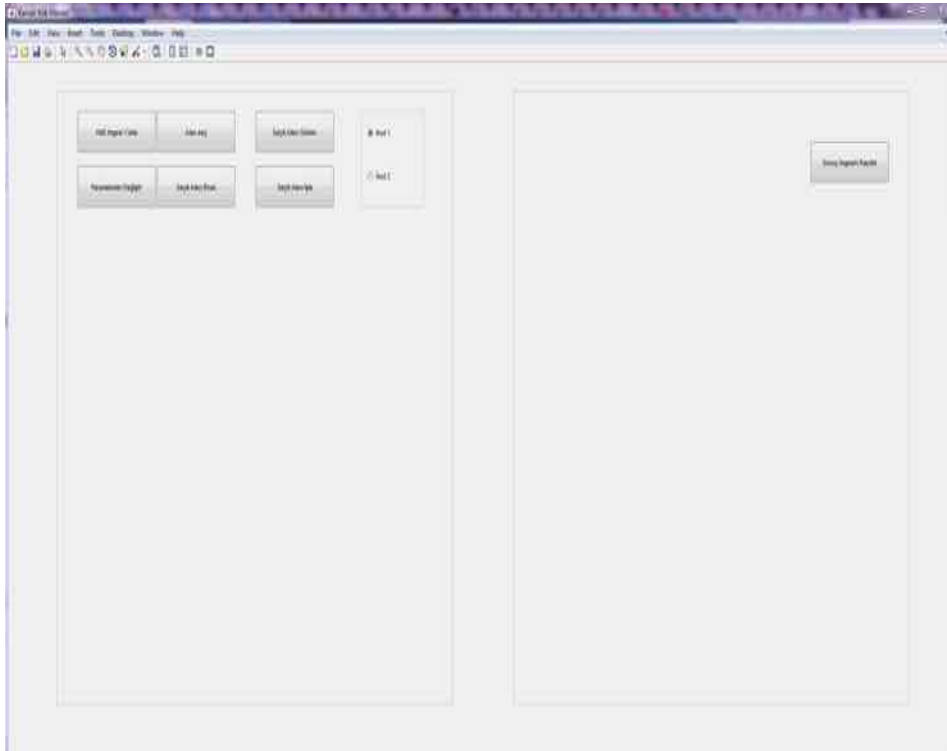
##### **Kullanım Kılavuzu:**

## Giriş:

By kısım “CSCDetection” uygulamasının doğru ve kolay kullanımı için kullanıcılara yönelik bilgiler içermektedir.

## Ana Pencere:

Kullanıcı yazılımı çalıştırdığında ana pencere aşağıdaki gibi görünmektedir:



Ekranın sol üst bölgesindeki tuşların yakılaştırılmış görüntüsü:

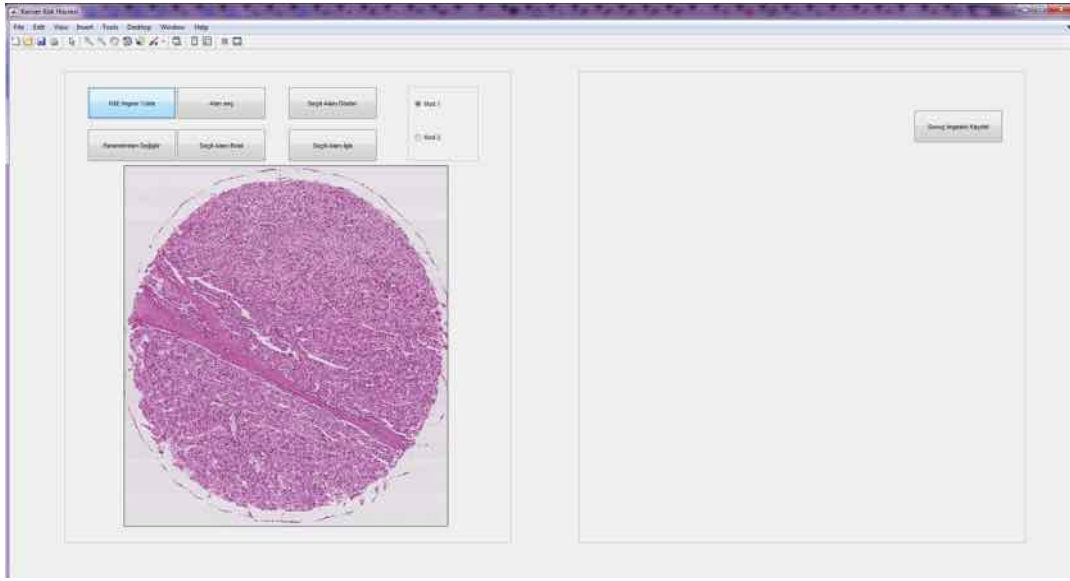


Bir imge yüklemek için “**H&E İmgesi Yükle**” tuşuna basın. Bununla birlikte, bir sılaytın bir kısmı ya da tamamı yazılıma yüklenebilmektedir. Burada önemli bir nokta, yazılımdaki algoritmaların düzgün çalışabilmeleri için yüklenen resmin bir RGB (kırmızı-yeşil-mavi) resmi

olması gerektir. Yazılım “.tif” ve “.jpeg” uzantılı resimleri kabul desteklemektedir.

Önemli bir nokta olarak, kullanıcı bir slaydın tamamını programda **işlememelidir**. Bunun sebebi, slaydlarda yırtılmış ya da boş kısımların olmasından kaynaklıdır. Bu sebeple kullanıcı yazılımda analiz etmek istediği bölgeyi dikkatli seçmelidir.

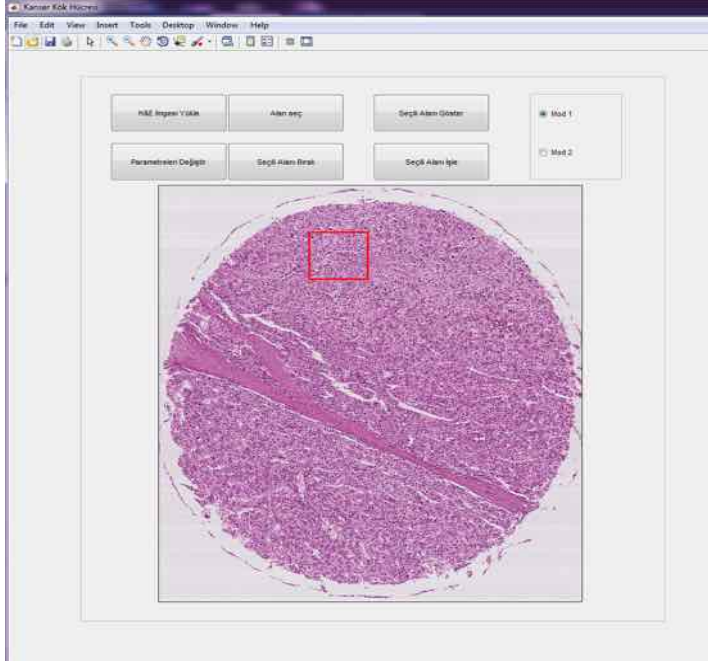
Bir imge yüklendiğinde pencerenin sol tarafında aşağıdaki gibi gözükcektir:



### İlgi Alanı Seçimi:

İlgi alanını seçmek için, “**Alan Seç**” tuşuna basın. Daha sonrasında ise hali hazırdaki açık imgeye tıklayarak ilgi alanını seçin. İmleci sürükleyerek bıraktığınızda ilgi alanını seçmiş olacaksınız. İlgi alanı seçildiğinde kırmızı bir kutu seçili alan etrafında oluşacaktır.



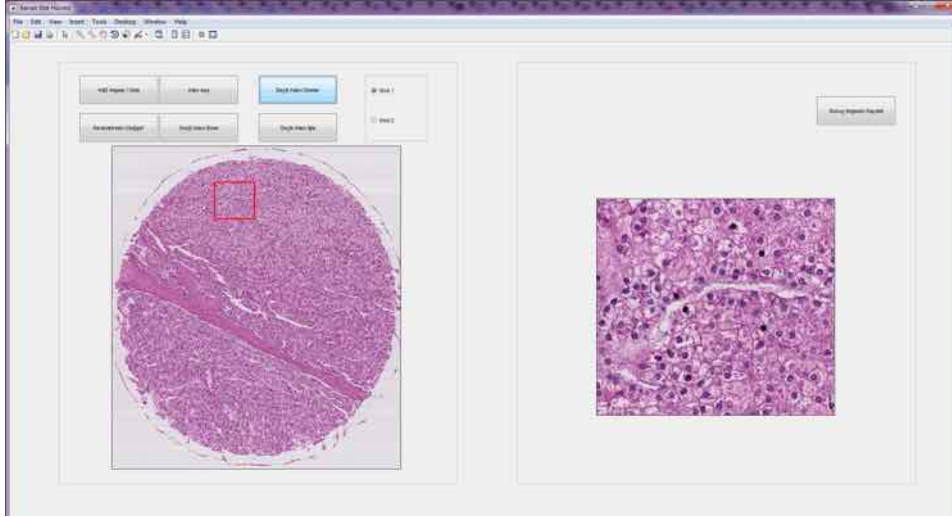


Seçiminizden memnun kalmazsanız ya da başka bir alan seçmek isterseniz aynı tuşa tekrar basın ve tekrar bir alan seçin. Yeni seçtiğiniz alan eskisinin yerine geçecektir. Seçimi silmek için **“Seçili Alanı Bırak”** tuşuna basın.

### **İlgi Alanının Gösterilmesi:**

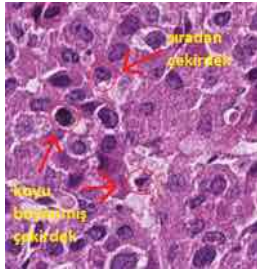
İlgi alanını işlemlerden geçirmeden önce **“Seçili Alanı Göster”** tuşuna basarak görebilirsiniz. Seçili ilgi alanı ana pencerenin sağ tarafında büyütülmüş olarak çizilecektir. Eğer hiç bir alan seçilmemişse ya da seçili alan silinmişse bütün resim sağ tarafa çizilecektir.





### Çekirdek Tanıma:

H&E ile boyanmış imgerlerde kanser kök hücrelerinin çekirdekleri, diğer hücrelerin çekirdeklerine göre daha koyu renktedirler. Bu durum aşağıdaki şekilde de görülebilmektedir:



Bu renk farkından yararlanarak kanser kök hücresi çekirdeği ve diğer hücrelerin çekirdeklerini birbirlerinden ayırmayı ve sınıflandırmayı hedeflenmektedir.

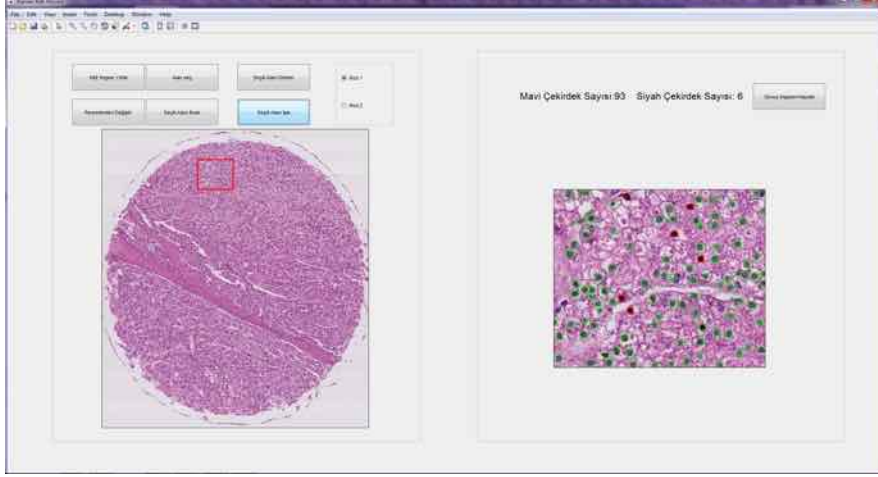
Bu amaçla yazılımın iki modu bulunmaktadır ve bu modlar iki seçenek düğmesiyle seçilebilmektedir.

2. mod kanser kök hücrelerinin çekirdekleri ve diğer çekirdekleri birbirlerinden ayırmak için bir yapay sinir ağı kullanmaktadır.

Bir mod seçtikten sonra **“Seçili Alanı İşle”**, tuşuna basın. İlgili alanının işlenmesi bittikten sonra sonuç resim ana pencerenin sağ kısmına çizilecektir. 2 modda da, sağ tarafta çıkacak işlenmiş imgede, kanser kök hücrelerinin çekirdekleri kırmızı bir çerçeve ve diğer çekirdekler ise yeşil bir çerçeve ile çizilmiş olacaktırlar.

Bununla birlikte toplam tanımlanmış kanser kök hücresi çekirdeği ve diğer hücrelerin çekirdeklerinin sayısı da pencerede görülebilecektir.

1 numaralı mod ile işlenmiş bir ilgi alanının örneği aşağıda görülebilmektedir:



2 numaralı mod ile işlenmiş bir ilgi alanının örneği aşağıda görülebilmektedir:



Unutmayın ki eğer bir ilgi alanı seçmezseniz bütün imge çekirdek tanımlaması işleminden geçecektir.

### **Yazılım Performansı:**

Performans testleri özellikleri aşağıda belirtilen bir bilgisayarda yapılmıştır:

#### **Donanım Özellikleri:**

- 1) Amd Phentom 560 2 Çekirdekli işlemcisi ~3.3 GHz.
- 2) 16-GB RAM

#### **Yazılım Özellikleri:**

- 1) Matlab 2016a
- 2) Windows 7 64-bit.

Aşağıdaki tabloda farklı büyüklüklerdeki ilgi alanları ve iki mod için işlem sürelerini içermektedir:

Mod\ imge büyüklüğü(piksel miktarı)	590*1460	989*1030
Mod 1	10.5 sn	13.9 sn
Mod 2	6.97 sn	6.97 sn

2 numaralı modun 1'e göre genel olarak daha hızlı olduğu gözlemlenmiştir.

### **İnce Ayar Parametreleri (1 Numaralı Mod İçin)**

Kullanıcı 2 parametreyi değiştirebilmektedir: “**Maksimum Çekirdek Büyüklüğü**” ve “**Siyah-Mavi Yağunluk Eşiği**”.

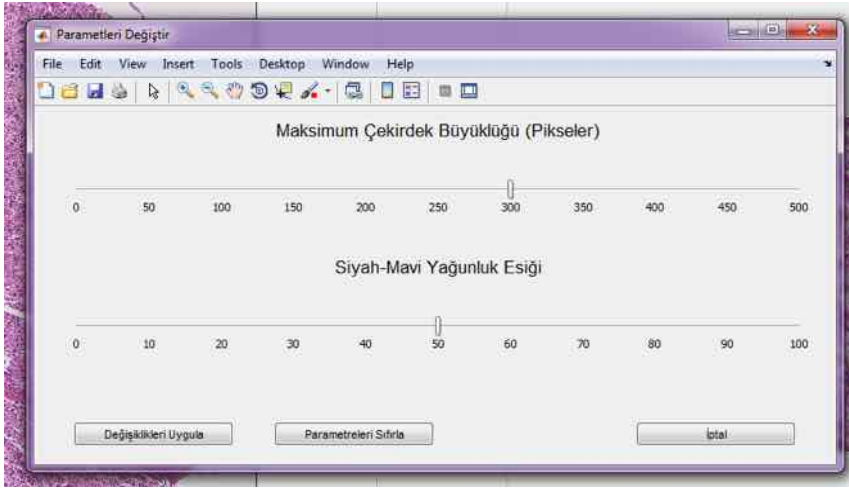
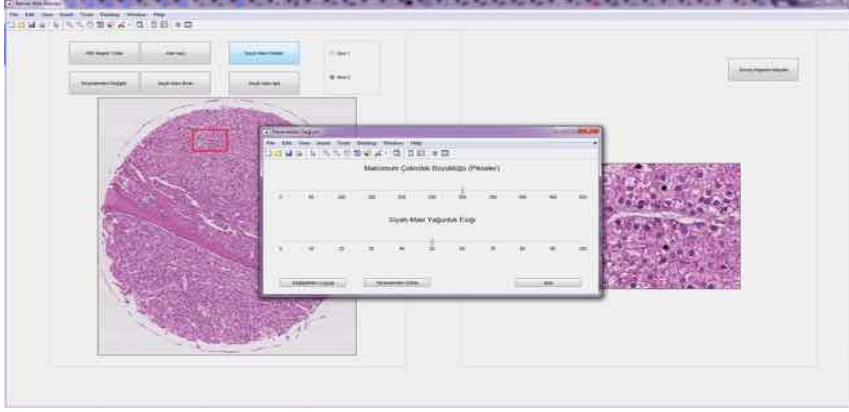
İlk parametre herhangi bir hücre çekirdeğinin resimde yaklaşık olarak kaç piksellik bir alan kapladığını belirtmek içindir. Eğer bir çekirdek bu alandan daha fazla alana yayılmışsa bu durumda belirlenmiş alanda birden fazla hücre çekirdeği olup olmadığını bulabilmek için yazılım imge üzerinde başka işlemler uygulayacaktır. Varsayılan değer 300 pikseldir.

İkinci parametre ise mavi ve siyah çekirdekler arasındaki tanımlama farkını belirlemek için kullanılan bir değerdir. İmgeler işlem öncesinde siyah-beyaz uzaya geçirildiği için piksel değerleri 0-255 arasında olacaktır. Bu durumda kanser kök hücrelerinin çekirdekleri diğer çekirdeklere göre siyaha daha yakın olacaktır ve düşük bir sınır değeri seçmek performansı artıracaktır. Varsayılan değer 60'tır.

Kullanıcı performansı artırmak için bu değerlerde değişiklikler yapabilmektedir.

### **Parametreleri Değiştirme:**

Yukarıda bahsi geçen parametreleri değiştirmek için “**Parametreleri Değiştir**” tuşuna basın. Yeni çıkan pencere aşağıdaki gibi olacaktır:



Değiştirilmiş parametreleri varsayılan hale getirmek için **“Parametreleri Sıfırla”** tuşuna basın. Yapılan değişiklikleri kaydetmek için ise **“Değişiklikleri Uygula”** tuşuna basın.

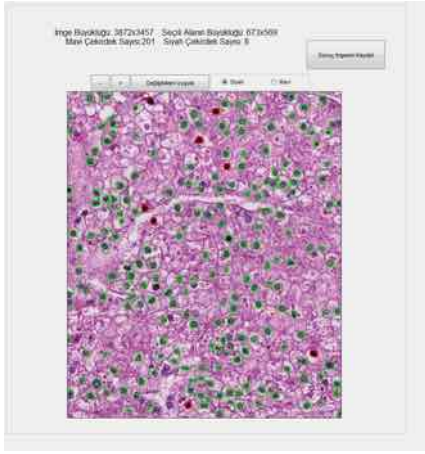
### **Sonuçları tekrar gözden geçirilmesi**

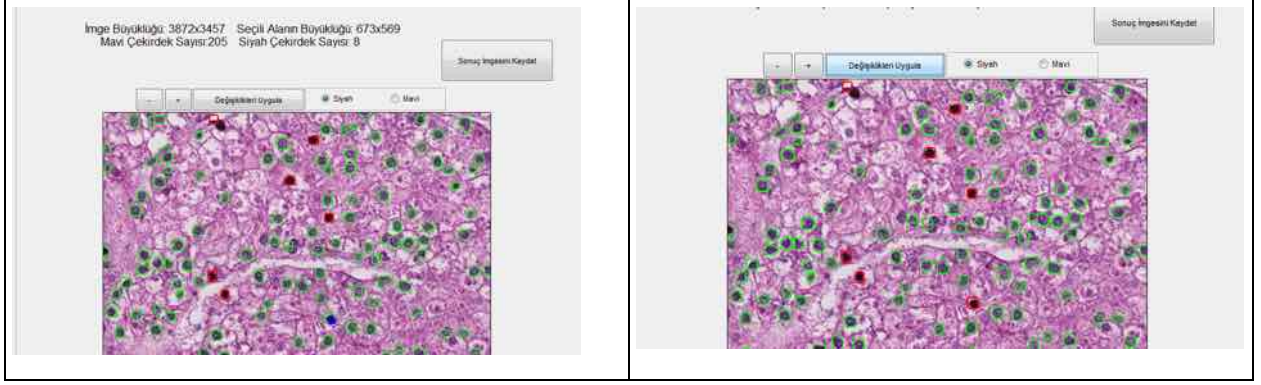
Yazılım bazı hücre çekirdeklerini yanlış sınıflandırabilir ya da hiç sınıflandıramaya da bilir. Güvenirliği artırmak ve garantilemek için kullanıcı sonuçları değiştirebilir, hücre çekirdeklerini programın arayüzünü kullanarak seçebilir, hatalı sınıflandırılmış çekirdeklerin sınıflarını değiştirebilir ya da çekirdek sınıflandırmalarını silebilir ve böylece yazılımın başarısını artırabilir. Bu değişiklikleri yapmak için kullanıcı, sonuç resminin üzerindeki seçenek düğmelerini kullanabilir.

Sınıflandırılmamış çekirdekleri sınıflandırmak için, radyo düğmelerinden çekirdeğin sınıfını seçtikten sonra **“+”** tuşuna basarak ve ardından da imge üzerindeki çekirdek bölgesine tıklamak yeterli olacaktır. Sonuç olarak istenilen çekirdek bölgesine mavi bir kutu çizilecektir. Yanlış sınıflandırılmış bir hücre çekirdeği için **“-”** tuşuna ve daha sonra da imge üzerinden yanlış sınıflandırılmış çekirdeğin üzerine tıklamak yeterli olacaktır. Bu işlem sonunda da imge üzerinde çekirdek etrafına sarı bir kutucuk çizilecektir. Gerekli değişiklikler yapıldıktan sonra

“Değişiklikleri Uygula” tuşuna basıldığında imge modifiye edilecek ve sayısal değer imgenin üzerinde yazacaktır. Yanlış sınıflandırılmış (örnek olarak, mavi yerine siyah ve ya siyah yerine mavi) çekirdek durumunda ise, çekirdeğin sınıfı üst kısımdaki tuş grubundan seçildikten sonra “+” tuşuna basarak ve daha sonra da imge üzerinden yanlış sınıflandırılmış çekirdeğin üzerine tıklamak yeterli olacaktır. Gerekli değişiklikler yapıldıktan sonra “apply changes (değişiklikleri uygula)” tuşuna basıldığında imge modifiye edilecek ve hücre çekirdekleri yeniden sınıflandırılmış olacaktır.

Aşağıdaki şekiller süreci anlatmaktadır:





### **İşlenmiş İmgeyi Kaydetme:**

İşlenmiş bir ilgi alanı imgesini kaydetmek için sağ taraftaki **“Sonuç İmgesini Kaydet”** tuşuna basın.

## Cancer Stem Cells Detection Software

CANSTEM software is designed to determine cancer stem cell ratio in microscopic images of cancerous tissues.

Cancer is a complex disease characterized by uncontrolled proliferation and invasion of abnormal cells. For this reason, cancer tissue includes cell groups called side-populations with different characteristics. The cancer stem cell population, being one of these groups, is defined as responsible for the poor prognosis of cancer. The ratio of cancer stem cell population is one of the factors that determines the survival of patients.

This project supported by TÜBİTAK (213E032). CANSTEM software was developed in order to guide the the chemotherapeutic treatment of cancer for the use of pathologists

CANSTEM software was developed by Dr. Enis Çetin, Dr. Ayşegül Üner and Dr. Rengul Atalay's research groups.

### Installation Instructions:

Introduction:

This section serves to instruct the download and installation processes of "CSCDetection" software on windows and OSX operating systems.

Windows:

*If you have MATLAB (R2016a) or later versions:*

1- Download" CSCDetection\_eng.exe".

2-Run the application "CSCDetection\_eng .exe".

*If you don't have MATLAB (R2016a) or later versions:*

1-Download either the internet installer or the standalone MCR file.

2-Run the file and install the MCR.

3- Run the application "CSCDetection\_eng .exe" from the CSCDetection's installation directory.

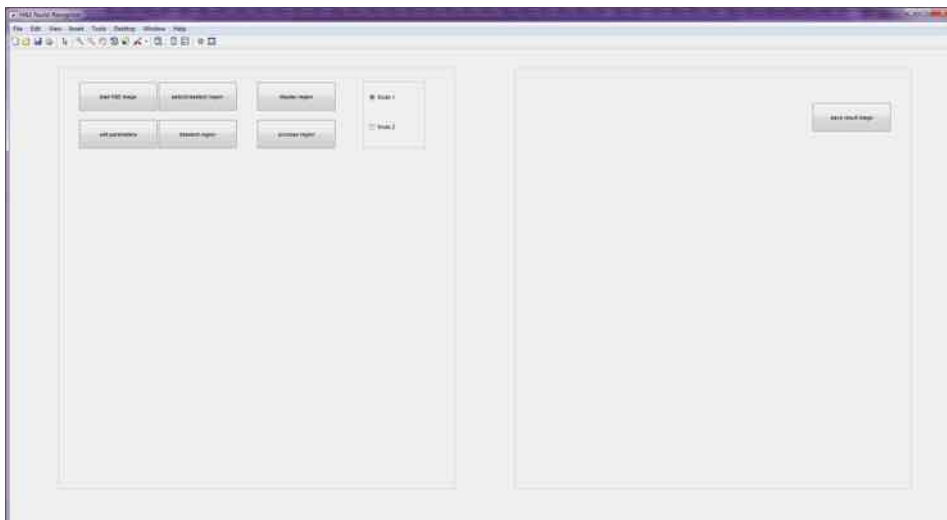
## Usage Instruction:

### Introduction:

This section guides the user as to how to use “CSCDetection” application properly and easily.

### Main Interface:

When the user opens the software, the main interface looks like the following:



These are the buttons that appear on the top right of the screen:



In order to load an image, **press “Load H&E image” button**. Notice that in order for the algorithm to run properly, the image should be an RGB image. In fact, one can load an entire slide or a portion of a slide.

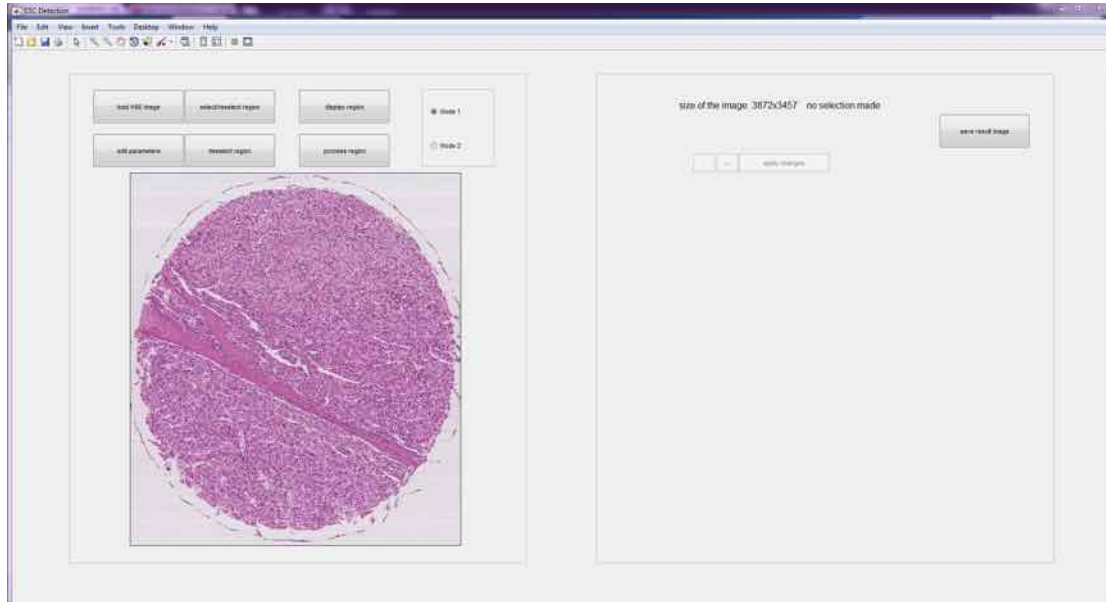
The software supports the following image extensions: tif, jpeg.

*Caution:* The user should not process an entire slide since some regions are torn and some parts may be useless. Therefore, the user should select a region of interest wisely and



cautiously( selection is explained in the next section).

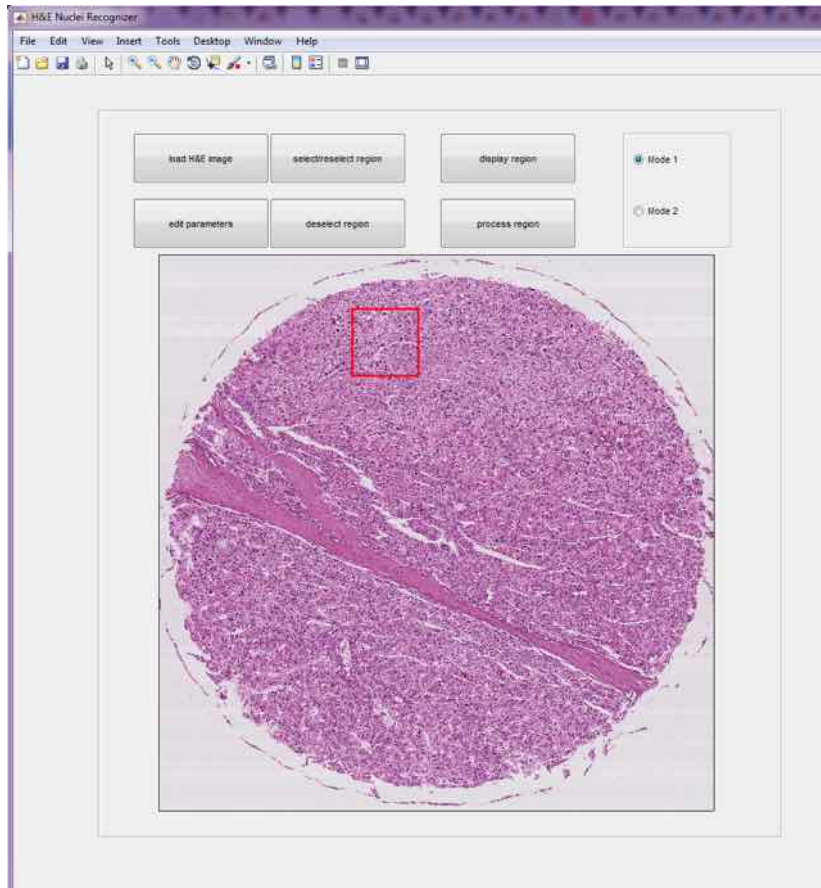
Once an image is loaded, it will appear on the left panel of the main interface.



### ROI Selection:

To select a region of interest, **press “select/reselect region” button**. Afterwards, click on the image to start selecting the region of interest. Keep dragging the cursor and release it eventually. Once the cursor has been released, a red rectangle will be drawn on top of the image to show the ROI.





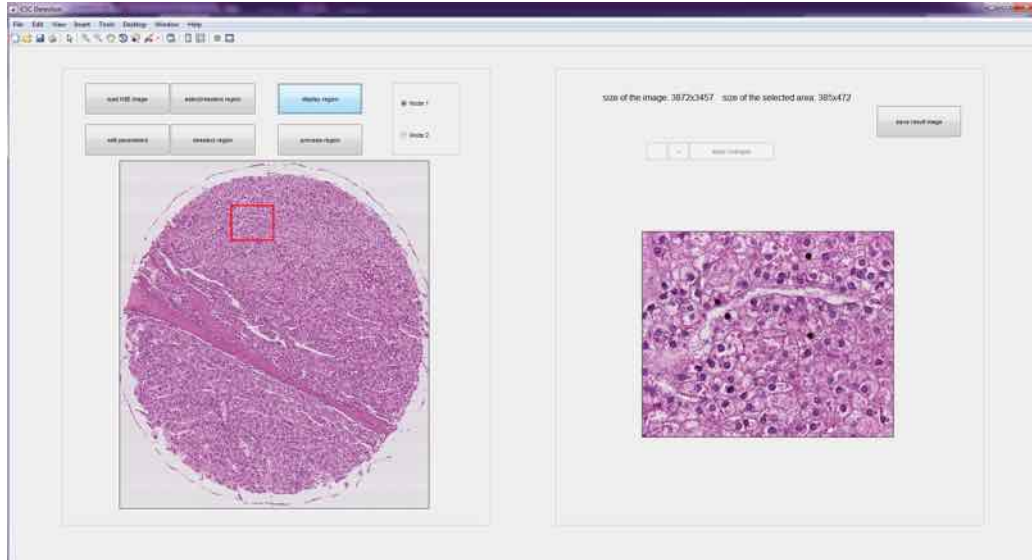
To re-select a region of interest, **press “select/reselect region”**. The new selection will overwrite the old selection.

To discard a selection, **press “deselect region”**.

### **Display Before Selection:**

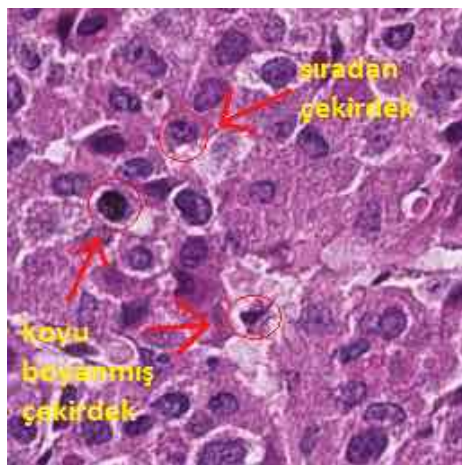
In order to view ROI before processing, **press “display region”**. The ROI will appear scaled up on the right panel. If no selection has been made or if discarded, the entire image will appear on the right panel.

The size of the selected area will appear on the top right of the interface.



### Nuclei Recognition:

In H&E image cancer-stem-cell nuclei appear darker than other nuclei as shown in the following figure:



Therefore, we aim at recognizing the nuclei and classify them into CSC nuclei and other nuclei.

In this regard, the software supports two modes for nuclei recognition. In order to choose one, **tick the desired mode in the radio button group.**

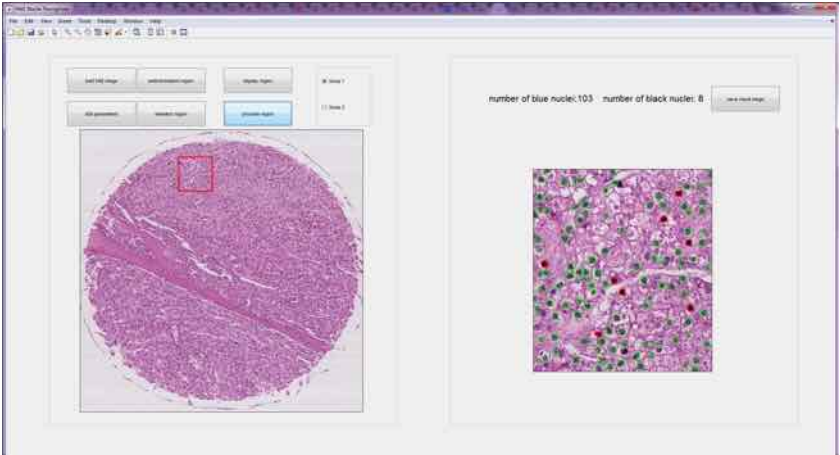
In fact, Mode 2 uses a neural network to classify CSC nuclei from the other nuclei while

After choosing a mode, **press “process region”**. Once the processing algorithm is done.

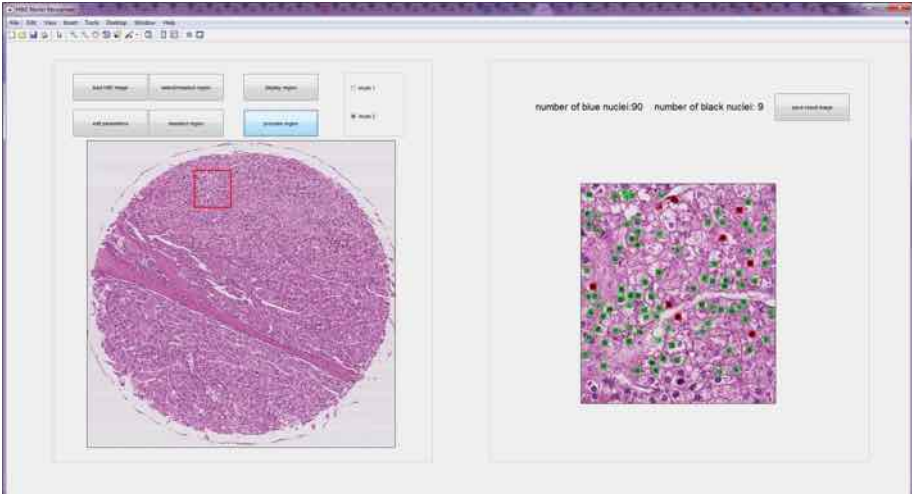
The resulting image will be drawn on the right panel. In the resulting image, if the image is processed by mode 1, the CSC nuclei will be circled by red contours and the other nuclei will be circled by green contours.

On the other hand, if the image is processed using mode 2, the CSC nuclei will be circled by red rectangles and the other nuclei will be circled by green rectangles. Moreover, the number of CSC nuclei and the the number of other nuclei will appear on the screen.

Here's an example resulting image for mode 1:



Here's an example resulting image for mode 2:



Notice that you can go back and forth between the selected image and the resulting image(s)

so long as the selection of the ROI is not changed.

Notice that if there's no selection, the entire image will undergo nuclei recognition.

### Software Performance:

Performance test was carried out on a machine with the following specifications:

#### Hardware specifications:

- 1) Amd Phentom 560 processor (2 CPU) ~3.3 GHz.
- 2) 16-GB RAM

#### Software specifications:

- 1) Matlab 2016a
- 2) Windows 7 64-bit.

The following table shows the elapsed time for processing ROIs with different sizes using the two modes:

Mode\ image size(in pixels)	590*1460	989*1030
Mode 1	10.5 seconds	13.9 seconds
Mode 2	6.97 seconds	6.97 seconds

Notice that mode 2 is generally faster than mode 1.

### Tuning Parameters (only applies for mode 1)

The user can tune 2 hyperparameters: “**Maximum Nucleus Size**” parameter and “**Black-Blue Intensity Threshold**” parameter to enhance the performance.

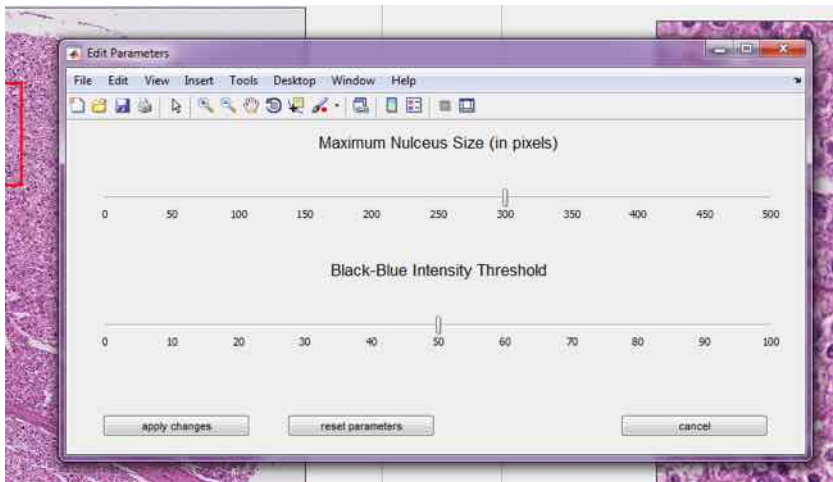
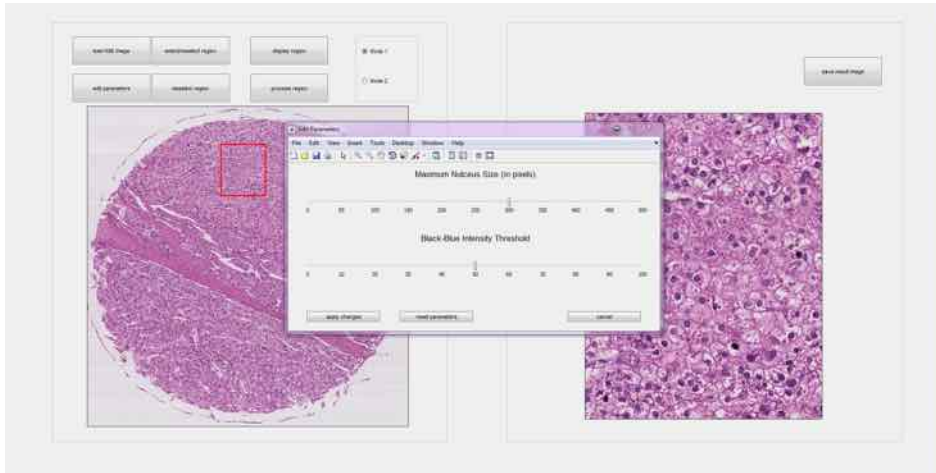
The first parameter is used to upper-bound the size of nucleus in pixels so that if an area exceeds this upper bound, it will undergo further processing in order to determine whether it has more than one Nucleus. The default value is 300 pixels.

The second parameter controls the boundary between blue and black nuclei based on the intensity value. Image intensity values are in between 0 and 255, which correspond to black and white colors, respectively. Indeed, a low value should be selected since CSC nuclei are dark. The default value is 60.

Notice that usually the user may need to do slight tuning, if any.

### How to edit parameters:

In order to edit the abovementioned two parameters, **press “edit parameter” button**. There, a new window will pop up as the following:



To reset the two parameters to their default values, **press “reset parameters” then “apply changes”**.

Notice that your choices will be saved during the session so long as you don't reset the parameters.

### Modifying Results:

The algorithm may miss some nuclei or misclassify them. In order to ensure reliability, the user can modify the result on the screen by adding, deleting or reclassify nuclei so that the resulting image can be as accurate as possible. In order to do so, choose the intended nuclei category from the radio button group right above the resulting image.

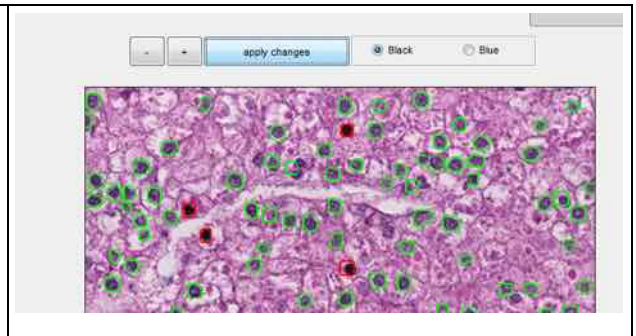
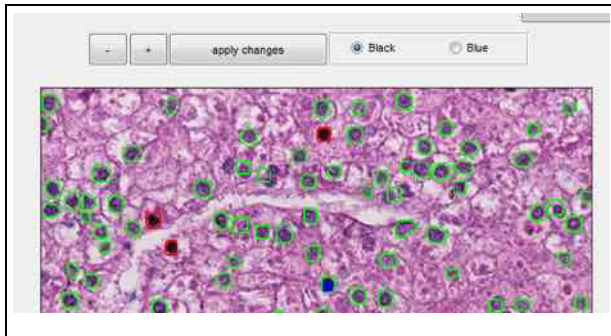
To add missing nuclei, **press (+)** and click on their locations on the resulting image. This will draw blue rectangles over these sites to indicate addition. To delete falsely recognized nuclei, **press (-)** and click on their locations. This will draw yellow rectangles over these sites to indicate deletion. Finally, **press “apply changes” button**. Then, the resulting image will be modified accordingly and so is the numeric result shown on top.

To re-classify a nucleus as black instead of blue or vice-versa, after specifying the intended category from the button group, **press (+)** and then click on the misclassified nucleus. **Press “apply changes” button**. Then, the nucleus will be re-classified without the need to delete it from the wrong class count.

Notice that this modification option works for both modes separately.

The following images illustrate the process:





### Save Resulting Image:

To save the resulting image that is currently displayed on the screen, press “save result image” on the right panel.





**TÜBİTAK**  
**PROJE ÖZET BİLGİ FORMU**

Proje Yürütücüsü:	Doç. Dr. RENGÜL ATALAY
Proje No:	213E032
Proje Başlığı:	Kanserli Dokuların Mikroskop Görüntülerinde Kanser Kök Hücre Oranın Otomatik Olarak Belirlenmesi ve Klinikte Kullanılacak Yazılım Geliştirilmesi
Proje Türü:	1003 - Öncelikli Alanlar (2. Aşama)
Proje Süresi:	24
Araştırmacılar:	ÇİĞDEM GÜNDÜZ DEMİR, AHMET ENİS ÇETİN
Danışmanlar:	AYŞEGÜL ÜNER
Projenin Yürütüldüğü Kuruluş ve Adresi:	ORTA DOĞU TEKNİK Ü.
Projenin Başlangıç ve Bitiş Tarihleri:	01/06/2014 - 01/12/2016
Onaylanan Bütçe:	564409.0
Harcanan Bütçe:	538196.25
Öz:	<p>Kanser anormal hücrelerin kontrolsüz çoğalması ve yayılması olarak tanımlanan karmaşık bir hastalıktır. Bu nedenle kanser dokusu farklı özellikleri olan-ek popülasyon denilen-hücre gruplarını içerir. Bu hücrelerden olan kanser kök hücreleri normal kök hücreleri gibi kendini yenileme ve farklılaşma özelliklerini taşıırken normal kök hücrelerinin aksine homeostatik kontrolleri olmayan, yani farklılaşma özelliklerini dengeleyip, çevresel sinyallere göre programlama özellikleri olmayan hücrelerdir. Bu özellikleri dolayısıyla da son yıllarda kanser kök hücre popülasyonu kanserin kötü prognozundan sorumlu hücreler olarak tanımlanmaktadır. Bu bağlamda kanser tedavisinde kanser kök hücre oran bilgisi, hastaya verilecek tedavinin planlanması amacıyla kullanılabilir. Hekimler kanserli dokunun kanser kök hücre popülasyon oranına göre henüz kanser metastaz yapmamış bile olsa bireye özgü tedavi yaklaşımı izlenebilir.</p> <p>Gerçekleştirilen bu proje kapsamında, kanserli parafin bloklara gömülü doku biyopsi örneklerinde, kanser kök hücre tanımlama öngörü aracını geliştirdik. kanserli hastalarda bireye ve hedefe yönelik tedavi yol haritasının belirlenmesine destek olabilecek ve hastalığın tedavisinde ve prognozunda iyileştirme sağlayabilecek yazılımı hekimlerin kullanıma sunduk. CANSTEM yazılıma <a href="http://users.metu.edu.tr/rengul/canstem.html">http://users.metu.edu.tr/rengul/canstem.html</a> adresinden ulaşılabilir.</p>
Anahtar Kelimeler:	Sayısal Patoloji, Kanser kök hücre, IHC, HxE, Görüntü analizi
Fikri Ürün Bildirim Formu Sunuldu Mu?:	Hayır