

Histoloji Görüntülerinde Kanserli Desenlerin Yarı GÜdümlü Öğrenme Yöntemiyle Tam Otomatik Sınıflandırılması

Automated Classification of Cancerous Textures in Histology Images Using Quasi-supervised Learning Algorithm

Devrim Önder¹, Sülen Sarioğlu², Bilge Karaçalı¹

1. Elektrik ve Elektronik Mühendisliği Bölümü,
İzmir Yüksek Teknoloji Enstitüsü
{devrimonder, bilgekaracali}@iyte.edu.tr

2. Tıbbi Patoloji Anabilim Dalı,
Dokuz Eylül Üniversitesi Tıp Fakültesi
sulen.sarioglu@deu.edu.tr

Özetçe

Bu çalışmanın amacı, histoloji kesitlerinden elde edilmiş görüntülerde yarı güdümlü istatistiksel öğrenme yöntemini kullanarak otomatik normal/kanserli desen sınıflandırmasını gerçekleştirmektir.

Bu çalışmada, insan kolon dokularına ait histoloji görüntüleri kullanılmıştır. İncelenen doku görüntüleri, normal ve kanserli durumlara ait görüntüler olmak üzere iki ayrı gruba ayrılmıştır. Her bir görüntü için eş-oluşum matrisleri kullanılarak desen öznelik vektörleri hesaplanmıştır. Normal ve kanserli gruplardan elde edilen desen öznelik vektörleri, yarı güdümlü istatistiksel öğrenme yöntemi ile sınıflandırılarak normal ve kanserli desen bölgeleri belirlenmiştir.

Terimler — Desen sınıflandırma, eş-oluşum matrisi, yarı güdümlü istatistiksel öğrenme.

Abstract

The aim of this work is to perform automated texture classification of histology slide images in health and cancerous conditions using quasi-supervised statistical learning method.

Tissue images were acquired from histological slides of human colon and were separated into two groups in terms of normal and disease conditions. Texture feature vectors corresponding to tissue segments of each image were calculated using co-occurrence matrices. Different texture regions were determined by the quasi-supervised statistical learning method using texture features of normal and cancerous groups.

Index Terms — Texture classification, co-occurrence matrix, quasi-supervised statistical learning.

1. Giriş

Literatürde, çeşitli alanlarda yapılmış desen analizi ve desen sınıflandırma çalışmaları mevcuttur. Yapılan çalışmalarda, genel olarak, belirli bir dokuda normal/kanserli desenlerin ayrıştırılması hedeflenmekte [1,2], ayrıca, uzmanlar tarafından uygulanan bazı doku desen derecelendirmelerinin otomatik olarak yapılabilmesi amaçlanmaktadır [3].

Desen sınıflandırma çalışmalarında çeşitli desen ölçütleri kullanılmaktadır. Sözkonusu çalışmalar kullandıkları desen ölçütlerine göre aşağıdaki bölümlere ayrılabilir;

1. Görüntü içerisindeki özel şekillere ait bilgilerin (hücre çekirdeği, kalsifikasyon nesnelerinin alanları, metrik boyutları vb.) kullanıldığı çalışmalar [4,5],
2. Doku görüntülerinin birinci ve ikinci dereceden istatistiksel özelliklerinin kullanıldığı çalışmalar (Haralick öznelikleri vb.) [6,7],
3. Dalgacık dönüşümlerinden elde edilen özneliklerin kullanıldığı çalışmalar [3].

Doku kesit görüntüleri kullanılarak hesaplanan desen öznelik vektörleri, k-en yakın komşuluk sınıflandırıcı, destek vektör makinesi, doğrusal ayrıştırma analizi, yapay sinir ağları gibi çeşitli sınıflandırma yöntemlerinden geçirilerek belirlenen sayıda (normal/kanserli vb.) sınıf içerisine alınarak analizler neticelendirilmektedir.

Bu çalışmada, doku kesitlerinin haritalandırılması için elde edilen her bir öznelik vektörünün sınıfının bilinmesini gerektirmeyen istatistiksel yarı güdümlü öğrenme yöntemi kullanılmıştır. Bu yöntem, kanser içerdiği bilinen bölgelerden çıkarılan desen vektörlerini, normal bölgelerden elde edilen vektörlerle kıyaslayarak normal ve kanserli desenleri otomatik olarak belirlemektedir.

Yapılan çalışmalar bir sonraki bölümde detaylandırılmıştır. Bölüm 3’de yöntemin, kolon histoloji görüntülerine uygulanmasıyla elde edilen sonuçlar gösterilmektedir.

2. Yaklaşım

2.1. Doku kesitlerinin görüntülenmesi

Yapılan çalışmada hematoksilin-eosin ile boyanmış kolon kesitlerinden sayısal mikroskop yardımıyla alınmış 1020x768 piksel boyutlu ve 4 mikron/piksel çözünürlükteki görüntüler kullanılmıştır.

2.2. Desen özneliklerinin hesaplanması

Bu çalışmada, histoloji görüntüleri üzerindeki bölgelerin

görsel özelliklerini belirleyen ikinci derece istatistiksel desen ölçütleri kullanılmıştır. Herhangi bir bölge için ayrı hesaplanan desen ölçütleri, çok boyutlu vektörler haline getirilerek, o bölgedeki doku yapısını karakterize eden bir desen profili oluşturulmuştur. Desen ölçütlerinin hesaplanmasında eş-oluşum matrislerinden yararlanılmıştır.

2.2.1. Eş-oluşum matrisleri

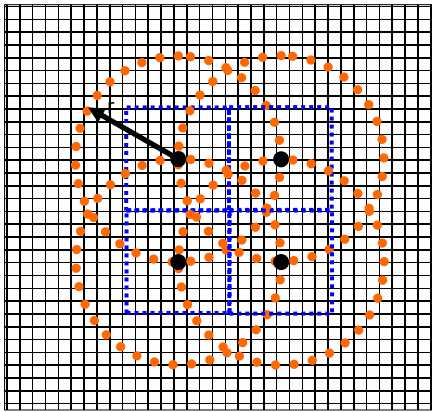
Gri seviyeli görüntülerde desen tanınmanın en bilinen yolu, belirli mesafelerdeki piksel çiftlerinde görülen gri seviyelerin eşli olasılık dağılımını yaklaşıklayan eş-oluşum matrislerinden türetilen desen ölçütlerini kullanmaktır [6].

Herhangi bir d pikseller arası mesafe için gri seviyeli bir I görüntüsünün x koordinat noktası merkezli ve r yarıçaplı $B_r(x)$ komşuluğundan elde edilen $M_{I,d}$ eş-oluşum matrisi

$$M_{I,d}(i,j) = \frac{\sum_{p,q \in Br(x)} I\{I(p) = i, I(q) = j, \rho(p,q) = d\}}{\sum_{p,q \in Br(x)} I\{\rho(p,q) = d\}} \quad (1)$$

ifadesi ile tanımlanır. Bu ifadede p ve q görüntüdeki pikselleri, $\rho(p,q)$ bu iki piksel arasındaki mesafeyi, $I\{*\}$ de argümanı doğru olduğunda 1, aksi halde 0 değerini alan bir fonksiyonu belirtmektedir.

Görüntü içerisindeki her bir piksel için yapılacak eş-oluşum matrisi hesabı için gereken toplam işlem süresi çok fazla olacağından, bir merkez nokta çevresindeki eş-oluşum matrisi değerlerinin, o yerel bölgedeki dağılıma yakınsadığı varsayımı yapılmıştır. Söz konusu merkez piksel çevresinden elde edilen öznelik vektörleri, o nokta merkezli disk içerisindeki kare şeklindeki bölgenin öznelik vektörleri olarak kabul edilmiştir. Şekil 1'de eş-oluşum matrislerinin hesaplama geometrisi gösterilmektedir.



Şekil 1. Eş-oluşum matrisleri hesaplama geometrisi. Koyu renkle gösterilen piksellere r yarıçapından daha yakın olan pikseller kullanılarak eş-oluşum matrisleri hesaplanmıştır.

2.2.2. Eş-oluşum matrisi tabanlı desen ölçütleri

Her bir d pikseller arası mesafe için aşağıda listelenmekte olan 14 öznelik değerinden oluşan bir vektör elde edilmektedir. Birden fazla d değeri kullanarak, her bir değer için hesaplanan 14 boyutlu vektörlerin uçuca eklenmesi ile bütün desen vektörleri elde edilmiştir.

1. Açısal ikinci kuvvet,
2. Kontrast,
3. Bağını,
4. Varyans kareler toplamı,
5. Ters fark kuvveti,
6. Toplam ortalama,
7. Toplam varyans,
8. Toplam entropi,
9. Entropi,
10. Fark varyansı,
11. Fark entropi,
12. Bağını bilgi özneliği 1,
13. Bağını bilgi özneliği 2,
14. En yüksek olasılık.

2.2.3. Desen özneliklerinin sıradüzenli biçimde hesaplanması

Sıradüzenli desen öznelikleri, belli bir merkez piksel etrafında, r^2 'nin sıralı katı olan farklı komşuluk yarıçapları ile hesaplanan eş-oluşum matrisi tabanlı desen ölçütlerinin uçuca eklenmesi ile elde edilir;

$$r = r_h, h = 1, 2, \dots, H. \quad (2)$$

Herhangi bir d pikseller arası mesafe için gri seviyeli bir I görüntüsünün x koordinat noktası merkezli ve r_h yarıçaplı $B_{r_h}(x)$ komşuluğundan elde edilen $M_{I,r_h,d}(i,j)$ eş-oluşum matrisi

$$M_{I,r_h,d}(i,j) = \frac{\sum_{p,q \in Br_h(x)} I\{I(p) = i, I(q) = j, \rho(p,q) = d\}}{\sum_{p,q \in Br_h(x)} I\{\rho(p,q) = d\}} \quad (3)$$

ifadesi ile tanımlanır.

2.3. Doku kesitleri üzerindeki normal ve kanserli bölgelerin yarı güdümlü istatistiksel öğrenme yöntemi ile belirlenmesi

Klinik durumlarla ilintili olan doku kesitleri, normal dokulardan alınan örneklerle kıyasladıklarında, görsel açıdan farklılıklar içermektedir. Hesaplanan desen ölçütleri, söz konusu değişik görsel karakterlerin belirlenmesini sağlamaktadır.

Doku görüntülerinde sağlıklı örneklerde farklı özellikler içeren bölgelerin saptanması ve tümörlü dokulardan elde edilen kesitlerde kanserli bölgelerin otomatik olarak işaretlenmesi uygulamalarında, güdümlü öğrenme teknikleri ağırlıklı olarak çalışılmıştır [1,8]. Güdümlü öğrenme yaklaşımlarında, hem normal hem de kanserli dokuların görüntü örneklerinin bir patoloji uzmanı tarafından elle işaretlenmesi, ardından da bu örneklerdeki dokusal yapıyı sayısal ölçütler kullanarak birbirinden ayıracak güdümlü sınıflandırma mekanizmalarının eğitilmeleri gereklidir. Bu şekilde, doku görüntülerindeki farklı yapıların (normal veya kanserli gibi) belirlenmesinde belli başarı oranları sağlanmış ise de, yöntemlerin uzman işaretlendirmelerine dayalı olması ciddi bir güçlük oluşturmaktadır. Ayrıca, güdümlü sınıflandırma teknikleri, kabul edilebilir başarı oranlarını yakalayabilmek için çok sayıda örneğe ihtiyaç duymaktadırlar. Doku görüntülerindeki farklı bölgelerin elle işaretlenmesi veya farklı dokuları örnekleyecek görüntü

kısımlarının ayıklanması ise son derece zahmetli ve zaman alıcı olduğu için, çok sayıdaki görüntünün bu şekilde belirlenip eğitimde kullanılmak üzere tanıma sistemine sunulması mümkün olmamaktadır. Bunun yanında, örnek olarak seçilen bölgelerin, işaretlemeyi yapan uzman veya uzmanların kişisel görüşlerine bağlı olması da kaçınılmazdır. Yetersiz sayıda örnek ve örneklerin saptanmasında kişisel görüşlerin baskınlığı bu yaklaşımla gerçekleştirilen görüntü analiz sistemlerinin hem ayrıştırma başarılarının hem de uygulama alanlarının kaçınılmaz olarak sınırlı kalmasına sebep olmaktadır.

Yapılan çalışmada, doku görüntülerindeki normal/kanserli bölgelerin belirlenmesi için yakın tarihte önerilen yarı güdümlü istatistiksel öğrenme yöntemi kullanılmıştır [4,9]. En yakın komşu sınıflandırıcısının özelliklerinden yararlanan bu yöntemin avantajı, eldeki görüntülerdeki doku bölgelerinin normal ve kanserli olarak elle işaretlenerek örneklendirilmesini gerektirmemesidir. Bu sayede, görüntülerden toplanan öznelik vektörleri yalnızca elde edildikleri doku kesitlerinde normal olmayan yapılar olup olmadığına bilgisine dayanarak gruplandırılmakta ve normal olmayan yapılara ait olanlar, normal görüntülerdekilerden istatistiksel ölçütler temelinde farklı vektörler olarak belirlenmektedir.

Yarı güdümlü öğrenme yöntemi, farklı iki sınıfa (normal veya kanserli) ait noktalar kümesi C_0 ve C_1 üzerinde çalışmaktadır. Bu durumda R kümesi, sözkonusu noktaların vektör değerleri ve noktaların ait olduğu sınıfların etiket değerlerinden oluşmaktadır:

$$R = \{x_i, y_i\}, y_i \in \{0,1\}, i=1,2,\dots, \ell.$$

Yarı güdümlü öğrenme yöntemi, R kümesinin, her iki sınıfa n tane nokta alınarak oluşturulmuş bir küme olduğunu varsayarak en yakın komşu sınıflandırıcısının sonucu x_i noktasının C_0 sınıfına ait olma olasılığını hesaplar. Benzer şekilde, x_i noktasının C_1 sınıfına ait olma olasılığı da hesaplanır:

$$\begin{aligned} f_0(x_i) &= Pr\{y=0\} \\ f_1(x_i) &= Pr\{y=1\} \end{aligned} \quad (4)$$

Yapılan çalışmada, kolon doku kesit görüntüleri normal ve kanserli olarak iki ayrı gruba ayrılmıştır. Bu sayede, hesaplanan tüm öznelik vektörleri normal veya kanserli olarak etiketlenmiş ve yarı güdümlü öğrenme yöntemine tabi tutulmuştur. Uygulama sonucunda her bir desen vektörüne ait $f_0(x)$ ve $f_1(x)$ değerleri elde edilmiş, bu değerler seviyelendirilerek sözkonusu vektörün normal veya kanserli bir desenle ilişkili olup olmadığına karar verilmeye çalışılmıştır.

3. Sonuçlar

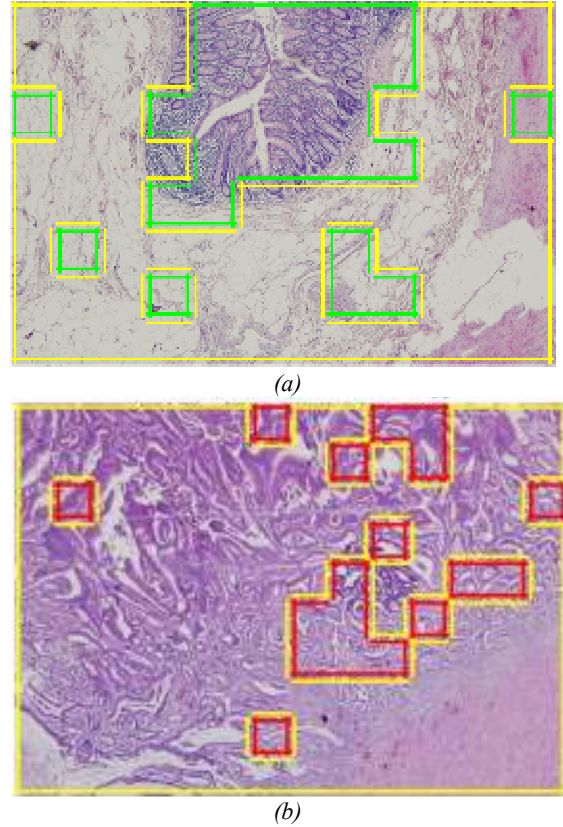
15 normal ve 15 kolon adenokarsinomlu görüntüyle oluşturulan deney grupları ile yarı güdümlü öğrenme yönteminin başarısı denenmiştir. Kullanılabilecek görüntü bilgileri (gri seviye veya renk bilgisi), öznelik hesaplama geometrileri (tek veya katlı r komşuluk yarıçapları) gibi değişkenlerin başarıma etkisini gözlemlemek için farklı deneyler düzenlenmiştir.

İlk deneyde, doku görüntüleri üzerinde, Şekil 1'de verilen geometri yardımıyla öznelik vektörleri hesaplanmıştır. Deneyde gri seviye bilgisi kullanılmıştır. Analiz yapılan diskin yarıçapı $r=64$ pikseldir. d olarak, sırasıyla 1, 3, 5, 9, 13, 17, 21, 41, 51 ve 61 değerleri

kullanılmıştır. Bu durumda, her bir bütün öznelik vektörü boyutu 140^2 'dir. Her bir vektör x_i için, hesaplanan n en uygun değeri ile $f_0(x_i)$ ve $f_1(x_i)$ olasılıkları hesaplanmıştır. Aşağıda verilen karşılaştırma kuralına göre, her bir x_i vektörünün ait olduğu desene y_i normal, kanserli ve bilinmeyen örneklere yapılmıştır;

$$\begin{aligned} f_0(x_i) - f_1(x_i) &\geq 0.9 \Rightarrow \text{normal}, \\ y_i = f_1(x_i) - f_0(x_i) &\geq 0.9 \Rightarrow \text{kanserli}, \\ |f_1(x_i) - f_0(x_i)| &< 0.9 \text{ durumlarda belirsiz.} \end{aligned} \quad (5)$$

Sözkonusu önermeler birbirleri ile birleştirilmiş bölgeler olarak farklı renklerle doku görüntüleri üzerine çizilmiştir. Yeşil renk normal, kırmızı renk kanserli ve sarı renk belirsiz desenlere karşılık gelmektedir. Şekil 2'de bu bölgelerin işaretlendiği kolon doku görüntülerinden iki örnek verilmiştir. Şekil 2.b'deki analiz sonucuna göre kanserli bölgelerin bir kısmı doğru sınıflandırılmış fakat oldukça fazla bilinmeyen olarak sınıflandırılmış kanserli bölge de mevcuttur.

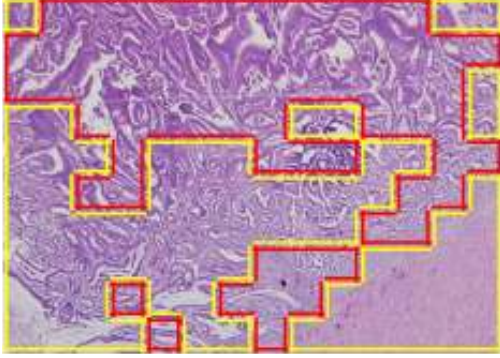


Şekil 2. Yarı güdümlü öğrenme yöntemi örnek deney sonuçları. (a) normal gruptan, (b) kanserli gruptan sınıflandırılmış bir görüntü. Bu deneyde gri seviye bilgisi kullanılmıştır.

Bir diğer deneyde, gri seviye bilginin yanında renk bilgisinin kullanılmasının sınıflandırma başarımına etkisini gözlemek amacıyla, Lab renk bilgisi kullanılarak elde edilen desen öznelik vektörleri ile yarı güdümlü öğrenme yöntemi yinelenmiştir. Desen öznelik vektörü olarak L, a ve b bileşenleri kullanılarak hesaplanan her bir tekil öznelik vektörü uçuca eklenerek elde edilen desen öznelik

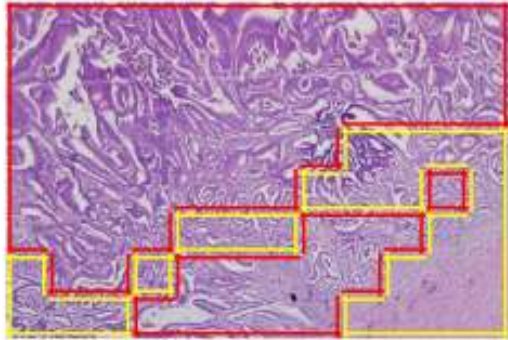
vektörleri kullanılmaktadır. Tüm değişkenler önceki deneyle aynıdır.

Renk bilgisinin kullanılması ile elde edilen sınıflandırma sonuçlarına ait örnek bir görüntü Şekil 3'de verilmiştir. Şekildeki histoloji görüntüsü Şekil 2.b'deki ile aynıdır. Renk bilgisi kullanılarak yapılan bu analizde, kanserli bölgelerin sınıflandırmasındaki başarının arttığı gözlenmektedir. Ayrıca, gerçekte kanserli olup bilinmeyen olarak ayrıştırılan bölgelerin sayısı da azalmıştır.



Şekil 3. Yarı güdümlü öğrenme yöntemi sonucunda elde edilmiş bir sınıflandırma görüntüsü. Bu deneyde Lab renk bilgisi kullanılmıştır.

Bir diğer deneyde ise önceki parametreler sabit tutulmuş fakat iki sıradüzenli öznitelik vektörleri hesaplanarak yarı güdümlü öğrenme yöntemi yinelenmiştir. Bu deneye ait bir örnek sonuç Şekil 4'de verilmiştir. Bu görüntüde ayrıştırılan kanserli bölgelerin Şekil 3'e göre daha yüksek ölçüde doğru olduğu görülmektedir. Buna paralel olarak, gerçekte kanserli olup bilinmeyen olarak ayrıştırılan bölgelerin sayısı da çok azalmıştır.



Şekil 4. Yarı güdümlü öğrenme yöntemi sonucunda elde edilmiş bir sınıflandırma görüntüsü. Bu deneyde Lab renk bilgisi kullanılmış ayrıca iki sıradüzenli eş-oluşum matrisleri kullanılmıştır.

4. Planlanan Çalışmalar

Yukarıda açıklanan çalışmanın devamında, kullanılan desen ölçütlerinden farklı ölçütlerin desen ayırıştırma başarısının denenmesi planlanmaktadır. Dalgacık dönüşümlerinden elde edilecek desen ölçütleri buna bir örnek olarak verilebilir. Desen sınıflandırma başarımının ölçülmesi için bir kontrol

deneyi ile birlikte başarımların ölçütlerinin tanımlanarak gözle yapılan başarımların değerlendirilmesinin bırakılması hedeflenmektedir. Bu sayede, yöntemde kullanılan desen ölçütlerinin ve diğer değişkenlerin etkinliği karşılaştırılabilecektir. Örnek olarak, en uygun r ve d değerleri belirlenebilecektir.

Gelecekte, bu çalışmada kullanılan iki sınıflı yarı güdümlü istatistiksel öğrenme yönteminin çok sınıflı şekilde genişletilerek normal/kanserli desen ayırıştırmasının yanında kanserli bölgelerin kendi içlerinde farklı desen sınıflarına ayrılması düşünülmektedir. Sonuç olarak, histoloji görüntülerinde klinik durumlara ilişkin farklı desen bölgelerini uzman müdahalesi olmadan tam otomatik olarak ayırıştırma sisteminin tasarımı ve geliştirilmesi hedeflenmektedir.

5. Kaynakça

- [1] Esgiar, A.N., et al., "Microscopic Image Analysis for Quantitative Measurement and Feature Identification of Normal and Cancerous Colonic Mucosa", *IEEE Transactions On Information Technology In Biomedicine*, 2(3): pp. 197-203, 1998.
- [2] Peter W. Hamilton, Peter H. Bartels, Deborah Thompson, Neil H. Anderson, Rodolfo Montironi and James M. Sloan, "Automated Location Of Dysplastic Fields In Colorectal Histology Using Image Texture Analysis", *Journal Of Pathology*, Vol. 182, pp. 68-75, 1997.
- [3] Kouros Jafari-Khouzani, Hamid Soltanian-Zadeh, "Multiwavelet Grading of Pathological Images of Prostate", *IEEE Transactions On Biomedical Engineering*, Vol. 50, No. 6, June 2003.
- [4] Karacali, B. and A. Tozeren, "Automated detection of regions of interest for tissue microarray experiments: an image texture analysis", *BMC Med Imaging*, 7: pp. 2, 2007.
- [5] A.P. Dhawan, Y. Chitre, C. Kaiser-Bonaso, "Analysis of mammographic microcalcifications using gray-level image structure features", *IEEE Trans. Med. Imag.*, 15 (3) pp. 246-259, 1996.
- [6] Haralick, R.M., "Statistical And Structural Approaches To Texture", *Proceedings Of The IEEE*, 67(5), pp. 786-804, 1979.
- [7] Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein, "Textural Features For Image Classification", *IEEE Transactions On Systems, Man, And Cybernetics*, vol. Smc-3, No. 6, November 1973.
- [8] Demir, C., S.H. Gultekin, and B. Yener, "Learning the topological properties of brain tumors", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2(3): pp. 262-270, 2005.
- [9] Karacali, B., "Quasi-Supervised Learning for Biomedical Data Analysis", *Pattern Recognition*, 2009. (in review).