

# Histopatolojik Görüntülerde Hücre Sınıflandırılması için İstatistiksel-Uzamsal Yaklaşım

## Statistical-Spatial Approach for Cell Classification in Histopathological Imagery

Mustafa Erseven<sup>1</sup>, Gökhan Bilgin<sup>2</sup>

<sup>1</sup>Elektronik ve Haberleşme Mühendisliği  
Yıldız Teknik Üniversitesi  
mustafaerseven@gmail.com

<sup>2</sup>Bilgisayar Mühendisliği  
Yıldız Teknik Üniversitesi  
gbilgin@yildiz.edu.tr

### Özetçe

*Bu çalışmada histopatolojik görüntülerin içerisinde yer alan etiketli hücrelerin sınıflandırılmasında uzamsal ilişkilerin etkileri değerlendirilmiştir. Öncelikle hücre merkezleri çevresinde farklı boyutlarda karesel pencereleyen hücresel yapıların Haralick, Tamura ve renk uzayları ile ilgili öznitelikleri çıkarılmıştır. Sonrasında ise 10 katlamalı çapraz doğrulama ile eğitim ve test veri setleri oluşturulmuştur. Ardından k-en yakın komşu, rastgele ormanlar ve destek vektör makinesi algoritmaları ile sınıflandırılmış, ve kullanılan bu sınıflandırıcılar üzerinde pencere boyutlarının değişiminin başarısı gözlemlenmiştir. Çalışmanın sonucunda en uygun pencere boyutu için bir yöntem tavsiye edilmiş ve karşılaştırma amacıyla elde edilen sonuçlar grafik ve tablolar şeklinde sunulmuştur.*

### Abstract

*In this paper, the effects of spatial relationships in the classification of labeled cells in histopathological images have been evaluated. Firstly, the features of the square windowing cells in different sizes related to Haralick, Tamura and color spaces have been extracted and have been merged. After that, the training and test data sets have been created by using 10 fold cross-validation. Then, data sets have been classified by k-nearest neighbors, random forest and support vector machine algorithms. The accuracies of different window sizes on these classifiers have been observed. As a conclusion, a method for optimal window size has been recommended and comparative results have been presented in the graphics and in the tables.*

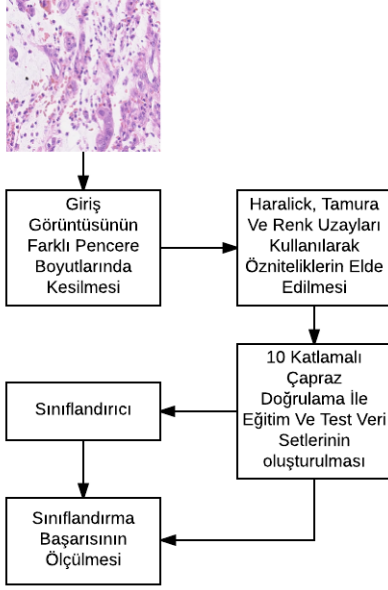
### 1. Giriş

Histopatolojik görüntülerin doku ile ilgili içerdiği bilgiden dolayı incelenmesi, hastalığın tespit edilmesinde, derecelendirilmesinde ve sınıflandırılmasında önemli bir

rol oynamaktadır. Doku içerisindeki hücrelerin incelenmesi ile erken kanser teşhisi konulması, söz konusu olan hastalığın tedavisi için kilit noktadır. Fakat görüntü sayısının fazlalığı, uzman kişiler tarafından ilgili görüntülerin incelenip karar verilmesi durumunda ciddi zaman ve insan gücü maliyetine neden olmaktadır. Bu probleme karşı, gelişen cihaz ve tekniklerin yardımı ile sayısal görüntülerin elde edilmesi ve bilgisarda işlenip, karar verilmesindeki ilerleme, dokuların analizinde görüntü işleminin önemini artırmıştır. Araştırmacılar için doku analizi ilgi çekici bir konu haline gelmiştir. H&E (hemotoksilen ve eosin) tekniği ile boyanmış dokular incelenirken dikkat edilen özellikler hücrelerin renkleri, şekilleri, dokusu gibi ayırt edicilerdir. Bu özelliklerin matematiksel modellere dökülebilir olması ve karar mekanizması olarak yapay öğrenmenin kullanılmasıyla başarılı sonuçların elde edilmesi, histopatolojik görüntülerinin incelenmesi işinde insan faktörünü önemli ölçüde azaltmıştır [9].

Önemli sayıda histopatolojik görüntülerde hücre sınıflandırmasıyla ilgilenmiş olan çalışma mevcuttur. Son yıllarda yapılan çalışmalarda, popüler olan derin öğrenme tekniklerinden mekansal olarak kısıtlı katlamalı sinir ağları (SC-CNN) hücrelerin sınıflandırılmasında ve tespit edilmesinde kullanılmıştır [1]. Ayrıca bir başka çalışma meme kanseri için histopatolojik görüntüleri derin öğrenme kullanarak sınıflandırmıştır [11]. Fisher vektör tanımlayıcı tabanlı yöntem ile histopatolojik görüntülerin kategorize edilmesi ve boyut azaltma ile sınıflandırma başarısının artırılmasıyla ilgili çalışılmıştır [12]. Histopatolojik görüntülerdeki hücrelerin doğası gereği, görüntü dokusu ayırt edici bir özellik olduğundan uzamsal yöntemler yoluyla öznitelikler elde edilmiş ve manifold öğrenme ile boyut azaltma yapılmıştır [2]. Benzer başka çalışmalar ise hücrelerin morfolojik özelliklerinden, hücre sayılarından ve konum bilgilerinden faydalanarak bulunan özniteliklerin kullanılması ile sınıflandırılmasıdır [3][6][13]. Sınıflandırma için kullanılacak özniteliklerin üzerine yoğunlaşan başka bir çalışma ise özellik çantası (bag of features) yaklaşımıdır. Elde ettiği özniteliklerin destek vektör makina-

larını kullanarak sınıflandırmıştır [10]. Görüntüleri farklı renk kanallarına ayırıp buradan elde ettiği özellikleri kullanarak yapılan çalışmalar da literatürde mevcuttur [4].



Şekil 1: Çalışmanın Akış Şeması

Bu çalışmada ise 100 adet 500x500 piksel boyutunda H&E ile boyanmış kolorektal adenokarsinomun histopatolojik görüntüsü kullanılmıştır (Bu veriler The University of Warwick - Department of Computer Science tarafından sunulmuştur). Her piksel 0,55 $\mu$ m/piksel (20xoptik büyütme) çözünürlüğündedir. Bu görüntüler 9 farklı hastadan 10 tüm görüntüden örtüşme olmaksızın kesilmiştir. Bu görüntülerin içerisinde sınıf etiketlerine sahip toplam 29756 hücre bulunmaktadır, bu hücrelerin 22444 tanesi sınıf etiketine sahiptir. Etiketli olan hücrelerin 20405 tanesi kullanılmıştır, 3 sınıftan oluşan bu örneklerden 7722 tane epithel, 5712 tane fibroblast ve 6971 tane hasta hücre mevcuttur [1]. Bu veri setini kullanarak sınıflandırma yapılması amaçlanmıştır. Görüntülerden doku, şekil ve renk özellikleri çıkarılarak, pencere boyutlarının (uzamsal ilişkisinin) sınıflandırma başarısındaki etkisi üzerine yoğunlaşmıştır. Tüm görüntüler üzerinden sınıf etiketli merkezlerden boyutları 17 x 17, 19 x 19, 21 x 21 ve 23 x 23 piksel olan pencereler kesilmiş, bu görüntülerde pek çok farklı özellik çıkarma yöntemi denenmiş bunlar içerinden en başarılı ve görüntünün farklı özellikleriyle ilgilenen Haralick, Tamura ve renk uzayları özellikleri birleştirilerek eğitim ve test kümeleri oluşturulmuştur. Gri seviye görüntüde her piksel için 4 farklı acıda bir piksel komşuluğa bakarak oluşturulan öz oluşum matrislerinden 56 tane Haralick özneliği, Tamuradan elde edilen 3 öznelik ve renkli görüntü üzerinden RGB ve HSV uzayından bulunan 12 öznelik, toplamda ise 71 adet öznelik kullanılmıştır. Ardından eğitici yöntemlerden k-en yakın komşu (K-EYK), rastgele ormanlar (RO) ve destek vektör makinesi (DVM) kullanılarak sınıflandırılmıştır.

## 2. Yöntem

Histopatolojik görüntülerden öznelik çıkarırken amaç ayırt edici noktaların iyi özümsemesidir. Görüntü üzerinden pencere içerisine aldığımız bölgede renk değişimleri ve dokunun homojen ya da heterojen dağılıyor olması, ilgili pencere içerisinde kalan hücrenin sınıflandırılması için gerekli ayırt edici bilgileri içerir. Hücrenin çevresiyle olan uzamsal ilişki de pencere boyutunu seçerken kritik olacaktır. Bu çalışmada pencere olarak 17 x 17, 19 x 19, 21 x 21 ve 23 x 23 piksel boyutunda kare matrisler kullanılmıştır. Ardından öznelikler çıkarılmıştır.

Haralick öznelikleri dokunun gri seviye eş oluşum matrisinden çıkardığı istatistiksel özellikleri kullanır. Eş oluşum matrisi piksellerin her birinin diğeri ile ilişkili olan konumuna ait bilgiyi taşır [5]. Bu dokunun uzamsal analizinde önemli bir rol oynar.  $K \times K$  boyutlarındaki eş oluşum matrisini karakterize etmek için kullanılan matematiksel denklemler tablo 1'de belirtilmiştir. Buradan 14 tane öznelik elde edilmiştir. Her pikselin 0, 45, 90 ve 135 derecedeki ilk komşularına bakarak oluşturulan 4 farklı eş oluşum matrisinden toplamda 56 tane Haralick özneliği elde edilmiştir [7][8]. Tamura özellikleri de doku ile ilgilidir. Gri seviye görüntüden 6 özellik çıkarmayı mümkün kılar. Bu özneliklerden 3 tanesi sınıflandırma için kullanılmıştır. Böylece gri seviye görüntüden dokuyla ilgili olarak 59 öznelik elde edilmiştir.

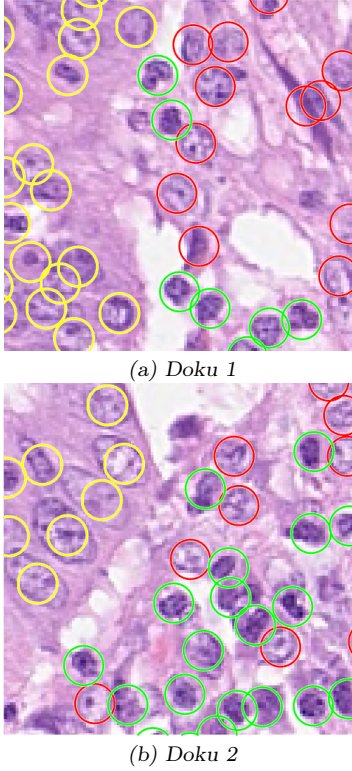
Ayrıca histopatolojik görüntülerin renkli olması ve bu bilgilerin hücreleri ayırt etmede gri seviye yöntemlerden farklı yeni özellikler kazandıracağı olmasından dolayı görüntüler farklı renk uzaylarında farklı renk kanallarına ayrılmıştır ve her renk kanalında ortalama ve standart sapması hesaplanmıştır. RGB uzayında kırmızı, yeşil ve mavi renk kanallarından 6, HSV uzayında renk özü, doygunluk ve parlaklık kanallarından da 6 öznelik olmak üzere toplamda 12 öznelikte renk ile ilgili olarak elde edilmiştir. Renkli görüntüsünde özneliklerinin kullanılması sınıflandırma başarısında etkili olmuştur. Böylece toplamda 71 öznelige sahip 20405 örnekten oluşan veri seti elde edilmiştir. 10 katlamalı çapraz doğrulama ile eğitim ve test veri setleri oluşturulmuş, K-EYK, RO ve DVM ile sınıflandırma yapılmıştır.

Tablo 1: Haralick Özelliklerinin Matematiksel Modeli

Açısal ikinci moment	$f_1 = \sum_{i=1}^K \sum_{j=1}^K p(i, j)^2$
Kontrast	$f_2 = \sum_{i=1}^K \sum_{j=1}^K (i - j)^2 p_{ij}$
İlinti	$f_3 = \frac{\sum_{i=1}^K \sum_{j=1}^K (i - \mu_i) p(i, j) - \mu_i \mu_j}{\sigma_i \sigma_j}$
Varyans	$f_4 = \sum_{i=1}^K \sum_{j=1}^K (i - \mu)^2 p(i, j)$
Ters fark momenti	$f_5 = \sum_{i=1}^K \sum_{j=1}^K \frac{1}{1 + (i - j)^2} p(i, j)$
Toplam ortalama	$f_6 = \sum_{i=2}^{2K} i p_{x+y}(i)$
Toplam varyans	$f_7 = \sum_{i=2}^{2K} (i - f_8)^2 p_{x+y}(i)$
Toplam entropi	$f_8 = - \sum_{i=2}^{2K} p_{x+y}(i) \log(p_{x+y}(i))$
Entropi	$f_9 = - \sum_{i=1}^K \sum_{j=1}^K p(i, j) \log(p(i, j))$
Varyans farkı	$f_{10} = \sum_{i=0}^{K-1} i^2 p_{x-y}(i)$
Entropi farkı	$f_{11} = - \sum_{i=0}^{K-1} p_{x-y}(i) \log(p_{x-y}(i))$
İlinti ölçümünün bilgisi 1	$f_{12} = \frac{H_{XY} - H_{XY}^1}{\max(H_X, H_Y)}$
İlinti ölçümünün bilgisi 2	$f_{13} = (1 - \exp(-2(H_{XY}^2 - H_{XY})))^{\frac{1}{2}}$ $H_{XY} = - \sum_{i=1}^K \sum_{j=1}^K p(i, j) \log(p(i, j))$
Maksimum ilinti katsayısı	$f_{14} = Q^{\frac{1}{2}}$ $Q(i, j) = \sum_k \frac{p(i, k) p(j, k)}{p_x(i) p_y(k)}$

### 3. Uygulama Sonuçları

RO, K-EYK ve DVM ile sınıflandırılmış farklı kare pencere boyutundaki veri setinin karşılaştırılması üzerinde durulmuştur. Daha öncede bahsedildiği gibi H&E ile boyanmış 100 histopatolojik görüntüde 3 sınıfa ait etiketleri bulunan 20405 hücre sınıflandırılmıştır. Bunlardan 7722 epitel, 5712 tanesi fibroblast ve 6971 tanesi hastalıklı hücredir. Şekil 2'de üzerinde çalışılan bir görüntü vermiştir. Sarı ile çembere alınanlar epitel, yeşiller hastalıklı ve kırmızılar ise fibroblast hücreleridir.

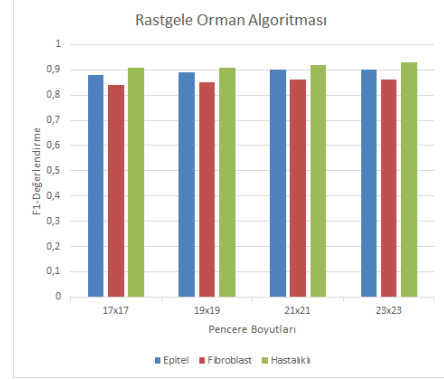


Şekil 2: Hücreleri İşaretlenmiş Histopatolojik Görüntüler

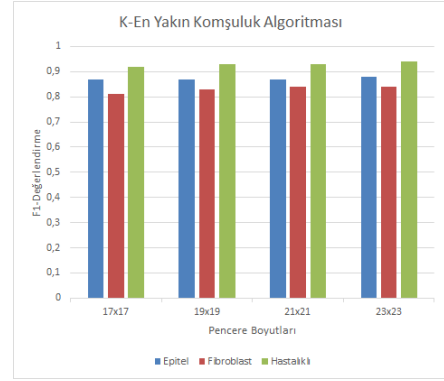
Merkez noktaları bilinen etiketli hücreler, merkezleri çevresinden farklı kare pencere boyutlarında ( $17 \times 17$ ,  $19 \times 19$ ,  $21 \times 21$  ve  $23 \times 23$ ) kesilerek RO, K-EYK ve DVM ile sınıflandırılmıştır. Sınıflandırıcıların farklı boyutlarda kesilmiş ve bu uzamsal boyutlardan çıkarılan öznitelikler (Haralick, Tamura ve Renk uzaylarından elde edilen 71 öznitelik) ile gösterdiği başarı karşılaştırılmıştır. 10 katlamalı çapraz doğrulama ile her sınıfın örnekleri rastgele eğitim ve test için ayrılmıştır. Örneğin epitel için 772 tane örnek test için, 6950 örnek ise eğitim için ayrılmıştır. Bu süreç 10 kez tekrarlanarak F1-ölçütü ve genel doğruluk hesaplanmıştır.

RO, K-EYK ve DVM algoritmaları Matlab'da gerçekleştirilmiştir. RO algoritması, sınıflandırma metodunda 45 ağaç ile ağaç yaprağı başına minimum gözlem sayısı 10 seçilerek uygulanmıştır[14]. k-en yakın komşu algoritmasının da ise 7 en yakın komşuya bakılarak uygulanmıştır. Destek Vektör makinelerinde ise LIBSVM kütüphanesi kullanılmıştır, 10-katlamalı çapraz doğrulama ile ya-

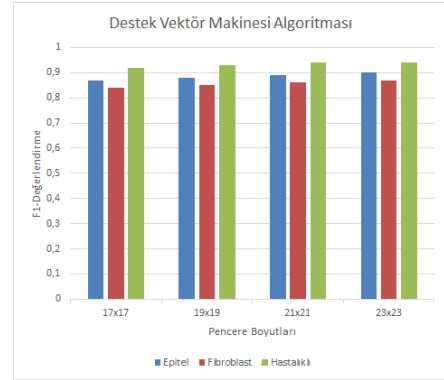
pılan parametre optimizasyonu sonucunda 'c' (cost) maliyet değişkeni 5 ve "g" (gamma) kernel fonksiyonundaki standart sapma değeri 0.1 olarak belirlenmiştir.



Şekil 3: RO Algoritmasının F1-Ölçütü Başarısı



Şekil 4: K-EYK Algoritmasının F1-Ölçütü Başarısı



Şekil 5: DVM Algoritmasının F1-Ölçütü Başarısı

Çalışmanın değerlendirilmesi Şekil 3, Şekil 4 ve Şekil 5'de F1-ölçütü, Tablo 2'de ise genel doğruluk şeklinde verilmiştir. Görüldüğü gibi 3 algoritma arasında büyük bir fark görünmemekle birlikte DVM algoritması daha başarılı olmuştur. Hücre merkezinin etrafından genişleterek kesilen görüntülerde görüldüğü gibi pencere boyutunun artması başarıyı artırmıştır. Bu uzamsal ilişki göz önüne alınırsa, pencere boyutu içerisinde hücrenin sınırları

## Kaynaklar

(mümkünse hücrenin tamamı ve farklı bir hücreyi içermek şartı ile) ve hücrenin etrafından aldığı pikseller ile çıkardığı özniteliklerin kullanılması yaklaşımı önerilmiştir. Çünkü, hücreleri sınıflara ayırma konusunda bu veriler gayet başarılı olmuştur. Pencere boyutunda kritik olan, mümkün oldukça diğer hücreleri pencereye dahil etmeden ilgilenen hücrenin sınırlarını ve etrafındaki piksellerin bir kısmını almaktır. Örneğin,  $11 \times 11$  kesilmiş görüntüler için epitel ve hastalıklı hücreler bu pencerenin içini tam doldururlar fakat fibroblast hücreleri bu pencere boyutundan çok daha büyük olduğu için pencere hücrenin iç kısmında kalır. Bu görüntüler sınıflandırılır ise fibroblast hücrelerinin başarısının düşük olduğunu görülmüştür (Örneğin RO algoritması ile  $11 \times 11$  boyutunda kesilmiş görüntüleri kullanırsak, fibroblast için F1-ölçütü 0,74'de, genel doğruluk ise 0,82'ye düşer).

Tablo 2: Uygulanan Sınıflandırıcıların Farklı Pencere Boyutlarına Göre Genel Doğruluğu

Pencere Boyutu	RO Gen. Doğ.	K-EYK Gen. Doğ.	DVM Gen. Doğ.
$17 \times 17$	0,88	0,87	0,88
$19 \times 19$	0,89	0,88	0,89
$21 \times 21$	0,90	0,88	0,90
$23 \times 23$	0,90	0,89	0,91

## 4. Sonuçlar

Bu çalışmada sayısal histopatolojik görüntülerin üzerinden doku, renk ve şekile dayalı çıkarılmış özniteliklerin farklı pencere boyutlarında 3 farklı algorithmada gösterdiği sınıflandırma başarısını gözlenmiştir. Pencere boyutunun en uygun büyüklüğünü bulmak için geliştirilen yaklaşım, bahsedildiği gibi pencerenin hücrenin sınırlarını ve çevresindeki piksellerin bir kısmını almasıdır. Çalışmamızda  $23 \times 23$  piksel boyutu en yüksek başarıyı göstermiştir. Ayrıca K-EYK, RO ve DVM algoritmaları içinde en yüksek başarıları yakalayan DVM olmuştur. Böylece histopatolojik görüntülerin içerisinde bulunan hücrelerin sınıflandırmasında uzamsal ilişkilerin etkisi üzerinde durulmuş bunun ışığında pencere boyutu seçmek için bir yaklaşım önerilmiştir.

Bu çalışmanın devamında, uyarlamalı pencereleme çalışılabilir. Hücre şekil ve konum olarak ön değerlendirilen ardından uygun pencere boyutuna alınır. Böylece farklı tür hücreler farklı boyutta pencerelenir. Elde edilecek özniteliklerin sınıflandırma başarısı, bu çalışma olduğu gibi tüm hücre çeşitleri için aynı boyutta pencere kullanan yöntemlerin başarısı ile karşılaştırılabilir.

## Teşekkür

Değerli arkadaşım Buğra Sofu'ya desteklerinden dolayı teşekkür ederim.

- [1] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.
- [2] D. Onder and B. Karacali, "Automated clustering of histology slide texture using co-occurrence based grayscale image features and manifold learning," in *14th National Biomedical Meeting, BIYOMUT'09*. IEEE, 2009, pp. 1-4.
- [3] M. Arif and N. Rajpoot, "Classification of potential nuclei in prostate histology images using shape manifold learning," in *International Conference on Machine Vision, ICMV'07*. IEEE, 2007, pp. 113–118.
- [4] J. Shi, J. Wu, Y. Li, Q. Zhang, and S. Ying, "Histopathological image classification with color pattern random binary hashing based pcanet and matrix-form classifier," *IEEE Journal of Biomedical and Health Informatics*, 2016.
- [5] R. C. Gonzalez and R. E. Woods, "Digital Image Processing." Prentice-Hall, 2002.
- [6] T. Dreyer, A. Doudkine, C. Popella, and C. Popella, "Nuclear shape and nuclear texture analysis in normal and neoplastic laryngeal epithelium," in *The First South-East European Symposium on Interdisciplinary Approaches in Fractal Analysis*, 2003, pp. 241–244.
- [7] R. M. Haralick, K. Shanmugam et al., "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 6, pp. 610–621, 1973.
- [8] E. Miyamoto and T. Merryman, "Fast calculation of haralick texture features," *Human computer interaction institute, Carnegie Mellon University, Pittsburgh, USA*. 2005.
- [9] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 97–114, 2014.
- [10] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, "Histopathology image classification using bag of features and kernel functions," in *Conference on Artificial Intelligence in Medicine in Europe*. Springer, 2009, pp. 126–135.
- [11] N. Bayramoglu, J. Kannala, and J. Heikkilä, "Deep learning for magnification independent breast cancer histopathology image classification," in *23rd International Conference on Pattern Recognition, ICPR'16*, IEEE, 2016, pp. 2440–2445.
- [12] Y. Song, Q. Li, H. Huang, D. Feng, M. Chen, and W. Cai, "Histopathology image categorization with discriminative dimension reduction of fisher vectors," in *European Conference on Computer Vision, ECCV'16* Springer, 2016, pp. 306–317.
- [13] F. Serin, M. Ertürkler, and M. Gül, "K-nearest unrepeatable cell graph model of histopathological tissue image," in *23th IEEE Signal Processing and Communications Applications Conference, SIU'15*. IEEE, 2015, pp. 2585–2588.
- [14] A. Liaw, M. Wiener et al., "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.