



Detection of Cervix Cancer from Pap-smear Images

 Fatma Betül AKYOL¹,  Oğuz ALTUN²

¹Corresponding Author; Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği;
<https://orcid.org/0000-0002-3836-1317>; akyolfbetul@gmail.com; +90 545 616 16 00

²Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği;
<https://orcid.org/0000-0001-8823-3148>; oaltun@yildiz.edu.tr

Received 18 April 2020; Revised 07 May 2020; Accepted 21 May 2020; Published online 28 August 2020

Abstract

Pap-smear test is used to detect cervical cancer, which ranks fourth in the ranking of cancer diseases in women worldwide. In this study, it is aimed to design a computer based decision system that can detect cervical cancer at an early stage. Normal and abnormal cells are found in the cervix images obtained as a result of the pap-smear test and the abnormal cells are marked on the image. The features extracted from the images were examined with pathologists and a dataset was created. For each of the 917 images in the Herlev dataset, these features were extracted and stored in a dataset. Support Vector Machines (SVM), Naive Bayes, Random Forest (RF), Multilayer Perceptron (MLP), Logistic Regression (LR), K- Nearest Neighbor (KNN) methods were applied to the created dataset, and accuracy values between 83% and 92% were obtained.

Keywords: Cervical cancer, Pap-smear, Segmentation, Classification

Pap-Smear Görüntülerinden Rahim Ağzı Kanseri Tespiti

Öz

Dünya çapında kadınlarda görülen kanser hastalıkları sıralamasında dördüncü sırada yer alan rahim ağzı kanserini erken aşamada tespit edebilmek için pap-smear testi kullanılmaktadır. Bu çalışma kapsamında rahim ağzı kanserini erken aşamada tespit edebilecek bilgisayar temelli bir karar sisteminin tasarlanması amaçlanmıştır. Pap-smear testi sonucunda elde edilmiş olan serviks görüntülerinde normal ve anormal özellikli hücreler bulunarak anormal olan hücreler görüntü üzerinde işaretlenmiştir. Görüntülerden çıkarılan özellikler patoloji uzmanları ile incelenmiş ve bir veri seti oluşturulmuştur. Herlev veri setinde bulunan 917 görüntünün her biri için bu özellikler çıkarılmış ve bir veri setine kaydedilmiştir. Oluşturulan veri setine makine öğrenmesi yöntemlerinden Support Vector Machines (SVM), Naive Bayes, Random Forest (RF), Multilayer Perceptron (MLP), Logistic Regression (LR), K- Nearest Neighbor (KNN) yöntemleri uygulanmıştır ve %83 ve %92 arasında doğruluk değerleri elde edilmiştir.

Anahtar Kelimeler Rahim ağzı kanseri, Pap-smear, Segmentasyon, Sınıflandırma

1. Giriş

Kanser, vücuttaki hücrelerin anormal büyümeleri nedeniyle meydana gelen ve günümüzde 100'den fazla çeşidi bulunan bir hastalık türüdür [1]. Bu kanser çeşitlerinden biri olan rahim ağzı kanseri kadınlarda en sık görülen dördüncü kanser türüdür. 2018 yılında dünya çapında yaklaşık olarak 570.000 kadında rahim ağzı kanseri görülmüş ve 311.365 kadın bu hastalık nedeniyle hayatını kaybetmiştir [2]. Servikal intraepitelyal neoplazi (SIN) anormallikleri ile ilişkilendirilen İnsan Papilloma Virüsü (IPV) rahim ağzı kanseri vakalarının %50'sinde hastalık kaynağı olarak gözlenmiştir [3]. Bethesda Sistemi SIN'i, düşük dereceli skuamöz intraepitelyal lezyon (DSIL), yüksek dereceli skuamöz intraepitelyal lezyon (YSIL) ve skuamöz hücre karsinoması olmak üzere üç gruba ayırır. Anormal görünümde olan fakat kanser hücresi olmayan kanser öncüsü olarak görülen hücrelere displazi denir. Bethesda Sistemi'ne göre yukarıda belirtilen hücre anormalliklerinden, DSIL hafif displazi, YSIL ise orta displazi olarak sınıflandırılabilir [4].

Hastalığın teşhisinde farklı yöntemler mevcuttur. Mevcut yöntemler arasında en yaygın olarak kullanılan tekniklerden biri görüntülemedir. Günümüzde rahim ağzı kanseri görüntüleme yöntemi olarak “Pap-Smear Testi” yaygın olarak kullanılmaktadır. [5].

Kanser hastalığında ve kanser kaynaklı ölümlerde meydana gelen artış bu hastalık üzerinde yoğun çalışmalar yapılmasını beraberinde getirmiştir. Yapay zekâ ile birlikte bilgisayar temelli karar sistemleri tasarlanmaya başlanmıştır ve bu karar sistemleri günümüzde tıp dâhil birçok alanda kullanılmaktadır. Makine öğrenmesi yöntemlerini kullanarak tıbbi görüntü analizi alanında yapılan çalışmalar yaygın şekilde devam etmektedir. Literatürde bu alanda yapılan birçok çalışma bulunmaktadır.

Taha vd. [6], pap-smear görüntülerinin analizinde özellik çıkarımı için Evrişimli Sinir Ağları (ESA), sınıflandırma için ise SVM yöntemini kullandılar.

Rahim ağzı kanseri teşhisinde serviks hücrelerinin çekirdek ve sitoplazmaları ayrı olarak ele alınmıştır. Bu nedenle hücreye çekirdek ve sitoplazma segmentasyonu işlemleri uygulanmıştır. Segmentasyon alanında yapılmış olan çalışmalardan, Plissiti vd. [7], pap-smear görüntülerinde hücre çekirdeği ve sitoplazmasını otomatik olarak sınıflandırmak için bulanık c-ortalama (fuzzy c-means) ve SVM tekniklerini kullandılar. Çalışmalarında hücre çekirdeklerini %99,25 doğruluk oranı ile sınıflandırdılar.

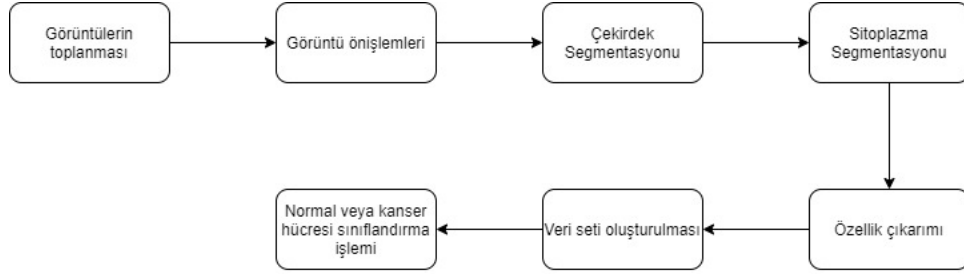
Wang vd. [8], hücre çekirdeğini bulma ve segmentasyonu için matematiksel morfolojiye sahip ortalama kaydırma kümeleme algoritmasını ve sınıflandırıcı olarak SVM yöntemini kullanmışlardır. Diğer bir çalışmada, Marinakis vd. [9], çalışmalarında özellik seçimi için genetik algoritmaları, sınıflandırma için KNN yöntemini kullanmışlardır. Genctav vd. [10], pap-smear görüntüleri üzerinde segmentasyon için watershed algoritmasını ve hücreleri sınıflandırmada ise denetimsiz öğrenme algoritmalarını kullanmışlardır. Phoulady vd. [11], genişletilmiş alan derinliği algoritmasını kullanarak serviks hücrelerinden çekirdek segmentasyonu işlemi yapmışlardır. %96 hassasiyet ile diğer metotlardan daha iyi bir sonuç elde etmişlerdir. Diğer bir çalışmada Win vd. [12] pap-smear görüntüleri üzerinde rahim ağzı kanseri tespiti için sınıflandırıcı olarak Linear Discriminant (LD), SVM, KNN, boosted trees (yükseltilmiş ağaçlar) ve bagged trees (torbalı ağaçlar) yöntemlerinin sonuçlarının kombinasyonu olan bir yöntem kullandılar ve bu yöntem ile iki sınıflı sınıflandırma işleminde %98,27 doğruluk değeri elde ettiler. William vd. [13] çalışmalarında pap-smear görüntülerinden serviks kanseri tespiti için bir araç geliştirme üzerine çalışma yapmışlardır. Çalışmalarında sınıflandırıcı olarak bulanık c-ortalama (fuzzy c-means) yöntemini kullanmışlardır ve %98,8 doğruluk değeri elde etmişlerdir.

Bu çalışma kapsamında pap-smear testi sonucunda elde edilmiş olan görüntülerde ilk olarak serviks hücrelerinin çekirdek ve sitoplazmaları ayrı olarak bulunmuştur. Çekirdek segmentasyonu için iki ayrı yöntem uygulanmıştır. Bu yöntemlerden ilki eşikleme (thresholding) yöntemi iken ikincisi K-ortalama kümeleme yöntemidir. Sitoplazma segmentasyonu için ise eşikleme yöntemi ve morfolojik işlemler kullanılmıştır. Segmentasyon işlemi ile görüntüde bulunan her bir çekirdek ve sitoplazmadan bazı özellikler çıkarılmıştır. Bu özelliklere hücrenin normal veya anormal olarak sınıflandırılmasında ihtiyaç duyulmuştur. Anormal olarak bulunan hücreler sınıflandırılmıştır ve görüntü hücre işaretlenmiştir. Görüntülerden çıkarılan özellikler bir veri setine kaydedilmiş olup patoloji uzmanları ile incelenmiştir. Çıkarılan özellikler bir pap-smear görüntü veri seti ile denenmiştir ve bir veri seti oluşturulmuştur. Oluşturulan veri seti SVM [14], RF [15], MLP [16], KNN [17], NaiveBayes [18] ve LR [19] yöntemleri kullanılarak normal ve anormal hücre sınıflandırılmasında kullanılmıştır.

Kullanılan materyal ve yöntemler 2. Kısımda anlatılmış olup, Sonuçlar 3. Kısımda belirtilmiştir.

2. Materyal ve Metotlar

Çalışmada verilerin elde edilmesi işleminden sınıflandırma işlemine kadar takip edilen adımlar Şekil 1’de gösterilmiştir.



Şekil 1 Çalışmada takip edilen adımlar

Çalışma segmentasyon, özellik çıkarımı ve sınıflandırma olmak üzere üç temel adımdan oluşmaktadır. Görüntüleri segmentasyon işlemine hazır hale getirmek için görüntü üzerinde bazı ön işlemler uygulanmıştır.

2.1 Veri Seti

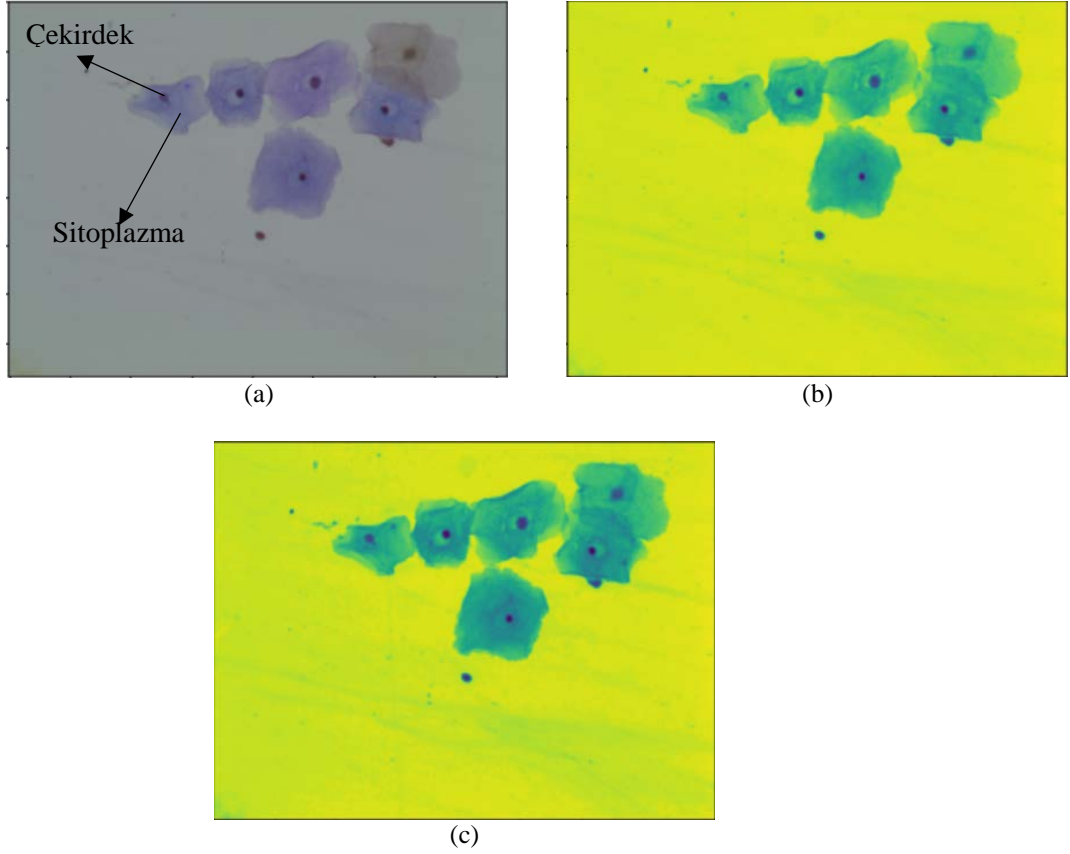
Bu çalışmada pap- smear görüntüleri kullanılmıştır. Çalışmada ele alınan yöntemler Herlev veri seti [20] üzerinde uygulanmıştır. Herlev Üniversitesi Hastanesi tarafından hazırlanmış olan Herlev veri seti 4 tanesi anormal ve 3 tanesi normal olarak belirlenmiş 7 sınıfa ait 917 pap-smear görüntüsünden oluşmaktadır. Anormal olarak belirlenen kategorilerden *carcinoma_in_situ*'ya ait 150, *light_dysplastic*'te 182, *moderate_dysplastic*'te 146 ve *severe_dysplastic*'te 192 pap-smear görüntüsü bulunmaktadır. Normal olarak sınıflandırılan kategorilerden *normal_superficial*'da 74, *normal_intermediate*'ta 70 ve *normal_columnar*'da 98 adet görüntü bulunmaktadır [21]. Görüntüler üzerinde sırasıyla ön işlemler, segmentasyon ve sınıflandırma işlemleri uygulanmıştır.

2.2 Veri Ön İşleme

Rahim ağzı kanseri tespitinde hücre çekirdeğinin önemli bir yeri vardır. Bu nedenle ilk olarak görüntü üzerinde hücre çekirdeği kısmı incelenmiştir. Çekirdeği tek olarak ele alabilmek için sitoplazma kısmı görüntüden çıkarılmıştır. Hücre çekirdeğini sitoplazmadan ayırabilmek için hücreye segmentasyon işlemi uygulanmıştır.

Bu çalışmada veri olarak pap- smear testi sonucu elde edilen serviks görüntüleri kullanılmıştır. Segmentasyon işlemine hazırlık için görüntüler bazı ön işlemlerden geçirilmiştir. İyileştirme işleminde amaç görüntü üzerinde bulunan gürültünün temizlenmesi ve kontrastlığın artırılmasıdır. Kontrastlığın artırılma amacı ise çekirdek ve sitoplazma arasındaki koyuluğun artırılarak çekirdeğin daha koyu sitoplazmanın daha açık hale getirilmesidir. Gürültüyü gidermek için boyutu 5 olan bir medyan filtre uygulanmıştır. Görüntüyü iyileştirmek için ise Ayrık Kosinüs dönüşümü (AKD) ve Ayrık Dalgacık Dönüşümü (ADD) yöntemleri uygulanmıştır.

Ayrık kosinüs dönüşümü görüntü sıkıştırma alanında kullanılan ve bu alanda yüksek performans gösteren bir yöntemdir [22]. AKD yönteminde amaç görüntüyü uzaysal alandan frekans alanına çevirerek iyileştirmektir. AKD yönteminde her pikselin grilik seviyelerinin tutulduğu 8×8 ' lik bir piksel matrisi oluşturulmuştur ve bu matrise AKD yöntemi uygulanmıştır. Girdi olarak verilen görüntüyü yeniden elde etmek için ters ayrık kosinüs dönüşümü uygulanmıştır. SIPAKMED [23] veri setinden alınan Şekil 2 (a)' da gösterilen giriş görüntüsü için AKD işlemi sonucu elde edilen çıkış görüntüsü Şekil 2 (b)'de gösterilmiştir.



Şekil 2 (a) Giriş görüntüsü (b) AKD sonrası görüntü (c) ADD sonrası görüntü

ADD, giriş olarak verilen görüntüyü ilk olarak L ve H olmak üzere 2 banda böler. 2 boyutlu ADD ise görüntüyü LL (low-low), LH (low-high), HL (high-low), HH (high-high) olmak üzere dört banda böler [24]. LH, HL ve HH bantları görüntünün yüksek frekansa sahip alanlarını içerirken LL ise düşük frekanslı alanını içerir [25].

Çekirdek ve sitoplazmayı birbirinden daha iyi ayırabilmek için bu iki görüntü arasındaki zıtlık oranı, çekirdek daha koyu, sitoplazma daha açık hale getirilerek artırılmıştır. ADD işlemi sonucu elde edilen görüntü Şekil 2 (c)' de de gösterilmiştir.

2.3. Çekirdek Segmentasyonu

Bir serviks hücrelerini normal veya anormal olarak sınıflandırabilmek için hücrenin çekirdeği ve sitoplazması ayrı olarak ele alınmıştır. Bu iki yapıyı birbirinden ayırabilmek için segmentasyon işlemi uygulanmıştır.

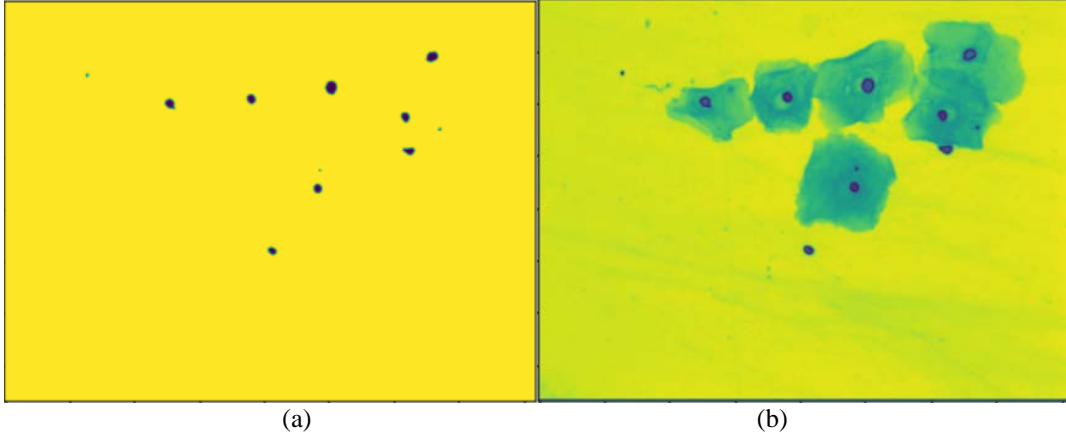
Şekil 2 (c)' de gösterilen ayırık dalgacık dönüşümü ile iyileştirilen görüntüde hücre çekirdeklerinin yaygın olarak mavi renk spektrumunda olduğu görülmektedir. Bu nedenle hücre çekirdeklerini tespit edebilmek için giriş görüntüsünde mavi renkli alanların bulunması gerekmiştir. Bu alanları bulabilmek için ilk olarak RGB renk uzayında bulunan giriş görüntüsü HSV (Hue, Saturation, Value) renk uzayına dönüştürülmüştür. Daha sonra mavi renkli alanları filtrelemek için Denklem 1'de a ile belirtilen mavi renk alt sınırı Denklem 2'de t ile gösterilen mavi renk üst sınırı belirlenmiştir ve görüntüde bu aralıklarda bir maske oluşturulmuştur. Alt ve üst sınır üç değerden oluşmaktadır Her bir değer HSV renk uzayının üç kanalına (h, s ve v) denk gelmektedir. H değeri rengi, s değeri rengin doygunluğunu ve v değeri rengin parlaklığını ifade eder. Renk uzayında bulunan tüm renkler belirli bir alt ve üst sınır renk değerleri arasında bulunmaktadır. a ve t dizilerinin ilk elemanları (0-120) renk uzayında mavi rengin alt(0) ve üst (120) renk değerini, ikinci elemanları rengin canlılık değerinin alt ve üst sınırını ve üçüncü elemanlar ise mavi renginin belirlenen parlaklık alt ve üst sınırlarıdır. Bu çalışmada mavi renkli bölgeler filtreleneceği için renk uzayında mavi renk alt ve üst sınırı [0,120] arası olarak belirlenmiştir. RGB renk

uzayında mavi renk $[0,0,255]$ şeklinde ifade edilmektedir. $[0,0,255]$ değeri HSV renk uzayına çevirilmiştir ve h değeri olarak 0 elde edilmiştir ve bu değer alt sınır olarak kullanılmıştır. Oluşturulan maske ile görüntüye Denklem 3'te belirtilen bit bazında mantıksal işlemler uygulanarak elde edilen m maske görüntüsü Şekil 3 (a)'da gösterildiği gibi maskelenmiştir.

$$a = [0, 0, 55] \quad (1)$$

$$t = [120, 255, 255] \quad (2)$$

$$g = f \& m \quad (3)$$



Şekil 3 (a) Maskeleye işlemi çıktısı (b) Hücre çekirdeklerinin giriş görüntüsü üzerinde işaretlenmesi

Maskeleye işlemi ile bulunan Şekil 2 (c) ile gösterilen f görüntüsü için sırasıyla morfolojik açma ve yayma işlemleri uygulanmıştır.

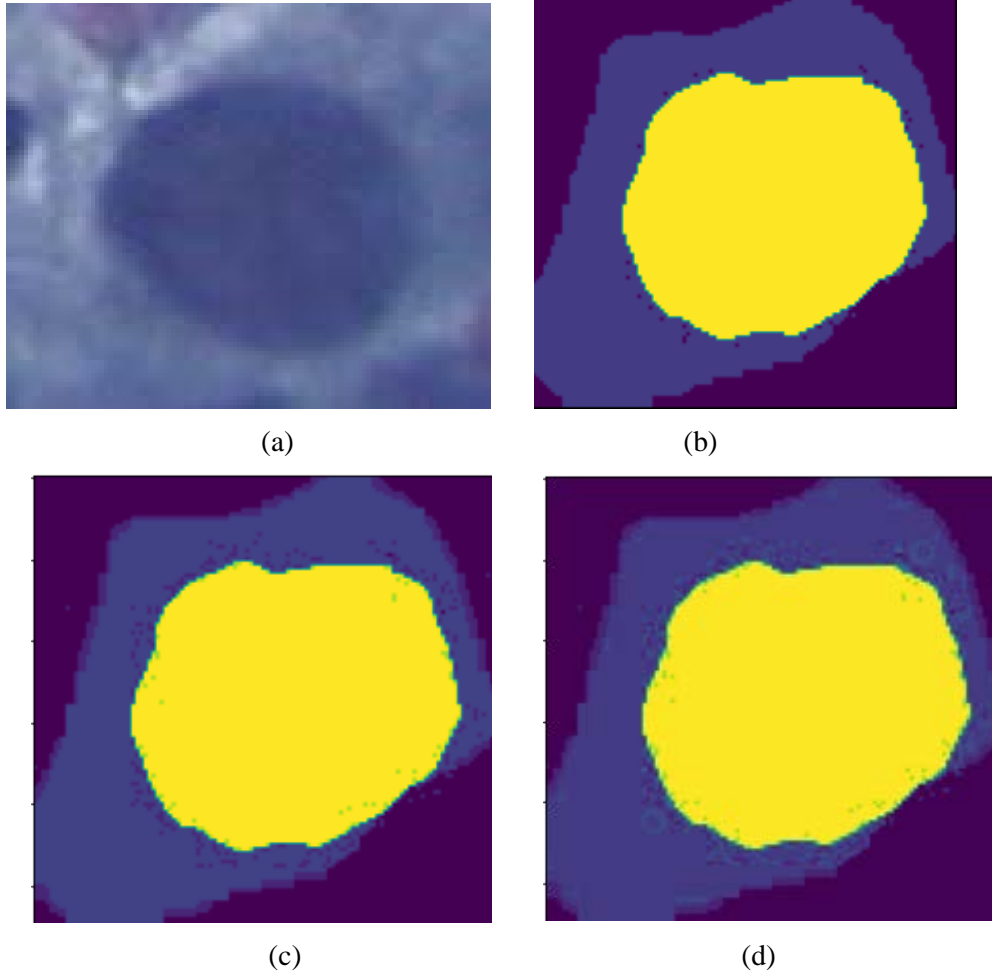
$$m = f \circ s \quad (4)$$

$$g = f \ominus s \quad (5)$$

Denklem 4'te HSV renk uzayında maske uygulanan f imgesi için s yapı elemanı ile uygulanan morfolojik açma işlemi belirtilmiştir. Denklem 5'te ise açma işlemi ve s yapı elemanı ile uygulanan morfolojik yayma işlemi verilmiştir.

Yayma işlemi sonucu elde edilen g görüntüsü Şekil 3 (a)'da gösterilmiştir. Elde edilen çekirdek maskeleri f giriş görüntüsü üzerinde Şekil 3(b)'de gösterildiği gibi işaretlenmiştir.

Çekirdek segmentasyonu işleminde kullanılan diğer bir yöntem ise K- ortalama (K-means) olmuştur. Bu yöntemde $k = 3$ alınmıştır ve Şekil 4 (a)'da verilen giriş görüntüsü için k-ortalama işlemi sonucu elde edilen görüntü Şekil 4 (b)'de gösterilmiştir. Farklı K değeri değiştirildiği zaman görüntü üzerinde yapılan kümeleme işlemi bazı noktaların yanlış sınıflara kümelendiği görülmüştür. Şekil 4 (c)'de görüldüğü gibi K-ortalama yönteminde $k = 5$ değeri kullanıldığında çekirdek içerisinde bulunan bazı noktaların yanlış kümelere atandığı gözlemlenmiştir. K değeri 7 olarak alındığında ise Şekil 4 (d)'de görüldüğü gibi kümeleme işlemi sonucu elde edilen görüntünün bozulduğu ve yanlış kümelere atanan noktaların sayısının arttığı gözlemlenmiştir.



Şekil 4 (a) Giriş görüntüsü (b) $k = 3$ için K-means sonucu (c) $k = 5$ için K-means işlemi sonucu (d) $k = 7$ için K-means işlemi sonucu

2.4. Sitoplazma Segmentasyonu

AKD ve ADD yöntemleri ile iyileştirilen smear görüntülerinde hücre ve sitoplazma arasındaki kontrast artırılmıştır. Çekirdek daha koyu, sitoplazma daha açık hale getirilerek segmentasyon aşamasında bu iki yapının birbirinden daha kolay ayrılabilmesi sağlanmıştır.

Hücrenin sitoplazma bölümlerini bulmak için uygulanan adımlar Denklem 6' da gösterilmiştir.

$$S_{stp} = G - f \quad (6)$$

Kontrastlığı artırılmış G giriş görüntüsünden eşikleme yöntemi kullanılarak bulunan çekirdek kısımları çıkarılmıştır. Kalan bölümlere OTSU eşikleme yöntemi ve Denklem 4 ve 5'te belirtilen morfolojik işlemler uygulanmıştır. Yapılan işlemler sonucu elde edilen görüntü Şekil 5'te gösterilmiştir. Görüntüde sitoplazma alanları beyaz olarak görünmektedir ve her biri kontur olarak kaydedilip özellik çıkarımı alanında belirtilen sınıflandırma kriterleri hesaplanmıştır.



Şekil 5 Segmentasyon işlemi sonucu sitoplazma

2.5. Özellik Çıkarımı

Bir serviks hücrelerini anormal veya normal olarak sınıflandırabilmek için görüntüden çıkarılan bazı özellikler sınıflandırma kriteri olarak kullanılmıştır. Anormal hücrelerde genellikle; konturda düzensizlik ve çekirdekte genişleme gözlemlenir. Koyu renk çekirdeğe sahip olup yüksek çekirdek-sitoplazma oranına sahiptirler [26].

Yukarıda belirtilen bilgilere dayanarak görüntü üzerinde aşağıdaki özellikler hesaplanmış ve sınıflandırma kriteri olarak belirlenmiştir:

Çekirdek alanı: Çekirdek olarak belirlenen her bir konturun alanı

Sitoplazma alanı: Sitoplazma olarak bulunan her konturun alanı

Çekirdek koyuluk değeri: Çekirdek konturlarının grilik seviyeleri

Çekirdek çevresi: Çekirdek konturlarının çevre hesabı

Sitoplazma çevresi: Sitoplazmaların hesaplanan çevre değeri

Çekirdeğin merkez koordinatları: Her bir çekirdeğin merkezinin bulunduğu (x, y) koordinatları

Sitoplazmanın merkez koordinatları: Her bir sitoplazmanın merkezinin bulunduğu (x, y) koordinatları

Çekirdek/ Sitoplazma (Ç/S) oranı: Hücrede bulunan çekirdek alanının sitoplazma alanına oranı

Çekirdek sitoplazma arası mesafe: Her bir hücre çekirdeğinin o hücrenin sitoplazması ile arasındaki uzaklık

Çekirdek yuvarlaklığı: Çekirdek konturunun yuvarlaklığının hesaplanması Denklem 7 ile gösterilmiştir.

$$Rnd = \frac{4 \cdot \pi \cdot alan}{cevre^2} \quad (7)$$

Sitoplazma yuvarlaklığı: Sitoplazmanın yuvarlaklık değerinin Denklem 7'ye göre hesaplanan değeri

Her hücrede bulunan çekirdek sayısı: Bazı anormal hücreler birden fazla çekirdekte oluşabildiği için sınıflandırma işleminde her bir hücrede kaç tane çekirdek bulunduğu hesaplanmıştır.

2.6. Sınıflandırma ve Değerlendirme

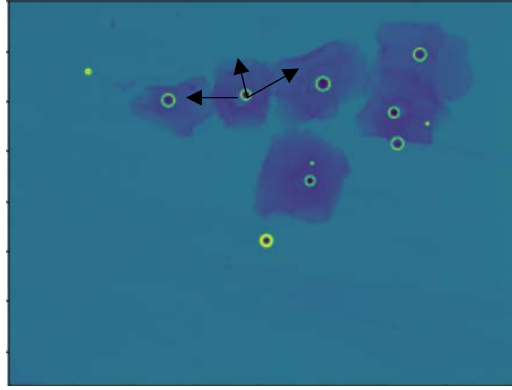
Çekirdek ve sitoplazma alanının hesaplanması çekirdek genişlemesi durumunun ölçülebilmesine olanak sağladığı için hücreyi anormal olarak sınıflandırmada büyük önem taşır. Bu nedenle her bir sitoplazma ve çekirdek konturu için görüntüden alan özelliği çıkarılmıştır. Elde edilen alan bilgilerinden önemli bir teşhis kriteri olan Ç/S oranı hesaplanmıştır. Segmentasyon işlemi sonucunda görüntüde birden fazla çekirdek ve sitoplazma konturları bulunduğu için her çekirdeğin içerisinde bulunduğu sitoplazması ile

eşleştirilmesi gerekmiştir. Bu eşleştirme işleminin doğru yapılmaması sonucunda yanlış Ç/S oranı hesaplanacağı için eşleştirme işlemi doğru oran bulmada önemli yer tutmaktadır. Eşleştirme işlemi aşamasında her bir çekirdek ve sitoplazma için çıkarılan merkez koordinatları kullanılmıştır.

Her bir çekirdek için (x_n, y_n) ikilisi ve her sitoplazma için (x_s, y_s) ikilisi konturun merkezinin bulunduğu x ve y koordinatlarını belirtmektedir. Her sitoplazma içinde çekirdek bulduđu için sitoplazmanın merkez koordinatları ile içinde bulduđu çekirdeğin merkez koordinatları arasında bulunan mesafenin az olması gerektiđi düşünölmüştür. Bu iki nokta arasındaki mesafeyi ölçmek için Denklem 8’te gösterilen Euclidean mesafesi yöntemi kullanılmıştır.

$$d = \sqrt{(x_n - x_s)^2 + (y_n - y_s)^2} \quad (8)$$

Şekil 6’da verilen referans görüntüsü için Denklem 8’te belirtilen d uzaklık değeri 1.00 değeri eşit veya daha küçük olması doğru çekirdek- sitoplazma eşleştirme sonucu verdiđi gözlemlenmiştir. d uzaklık değeri olarak hesaplama işleminde doğru eşleştirme sonucunu veren minimum değer kabul edilmiştir. Minimum değer alınmasının nedeni ise d mesafesinin yüksek bir değer alınması durumunda Şekil 6’da üzerinde örnek olarak alınan bir çekirdek oklar ile gösterilen üç sitoplazma ile de eşleştirilebilir ve bu durum yanlış çekirdek-sitoplazma eşleştirme sonucunu ortaya çıkarmaktadır. Sitoplazması ile eşleştirilen her bir çekirdek yuvarlak içine alınmıştır.

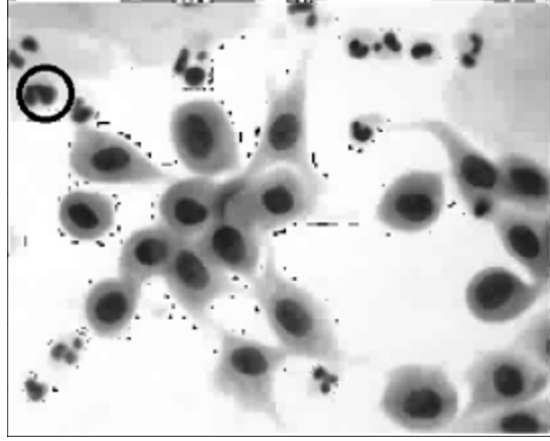


Şekil 6 Çekirdek-Sitoplazma eşleştirme

Eşleştirilen her bir çekirdek ve sitoplazma için Ç/S oranı hesaplanmıştır ve veri setine kaydedilmiştir. Veri setinde mesafesi 1.2 ve altında olan hücrelerde Ç/S oranı olarak hesaplanan değer, diğer durumlarda 0 olarak alınmıştır.

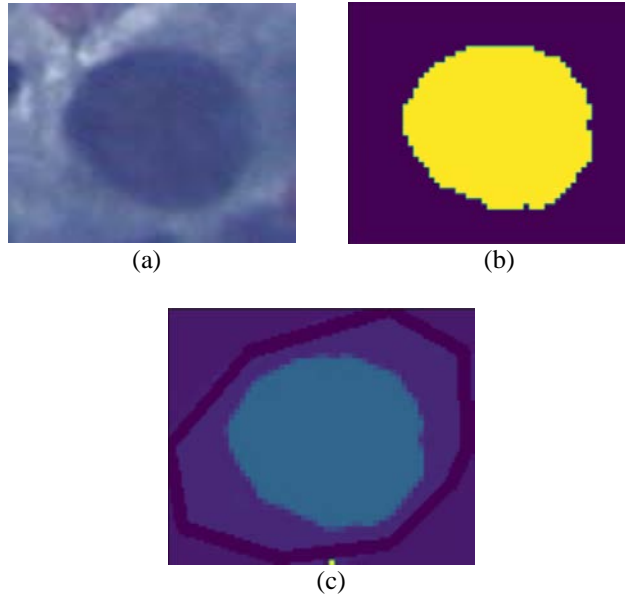
Koyu renk çekirdeđe sahip olan hücreleri bulmak için her bir çekirdek konturunun ortalama koyuluk değeri hesaplanmış ve ilgili çekirdek ile ilgili verilere kaydedilmiştir. Daha sonra veri setinde yoğunluk özelliğinin altında bulunan tüm değerlerin ortalaması bulunmuştur. Bir çekirdeğin konturu hesaplanan ortalama değeri büyükse hücre koyu renk çekirdeđe sahip olduđu düşünölmüştür.

Anormal hücre sınıflandırma kriterlerinden biri olan sitoplazmada bulunan hücre sayısı hesaplanmıştır. Eğer bir sitoplazmada birden fazla çekirdek bulunuyorsa bu hücreler anormal olarak değerlendirilmiştir. Bu hesaplama işleminde çekirdek konturlarının merkezlerinin arasındaki mesafe bulunmuştur. Mesafe hesaplama işleminde her (x_n, y_n) çekirdek merkez koordinatları için Denklem 8 ile verilen mesafe bulunmuştur ve bu d mesafesinin 5.00’in altında olması durumunda sitoplazma içerisinde bulunan çekirdek sayısı artırılmıştır. Şekil 7’de verilen görüntü [27] için birden fazla çekirdeđe sahip olan hücre bulunmuştur ve görüntü üzerinde yuvarlak içerisine alınmıştır.



Şekil 7 İki çekirdeğe sahip hücrenin görüntü üzerinde görünümü

Makine öğrenmesi yöntemleri ile başarı oranı yüksek bir sınıflandırma yapabilmek için veri sayısının fazla olması gerekmektedir. Bu nedenle Herlev veri seti kullanılarak bu veri setinde bulunan 917 görüntü için ilk olarak çekirdek ve sitoplazma segmentasyon işlemleri yapılmıştır. Şekil 8 (a)'da belirtilen görüntü için çekirdek ve sitoplazma segmentasyonu işlemlerinin sonucu Şekil 8 (b) ve (c)'de gösterilmiştir.



Şekil 8 (a) Giriş görüntüsü (b) Çekirdek segmentasyonu (c) Sitoplazma segmentasyonu

Çekirdek ve sitoplazma segmentasyonu işlemlerinden sonra sınıflandırma kriteri olarak belirlenmiş özellikler veri setinde bulunan 917 görüntü için bulunmuştur. Görüntülerden çıkarılan özellikler ile oluşturulan veri setinden normal ve anormal hücelere ait bazı veriler Tablo 1'de gösterilmiştir. Herlev veri setinde bulunan 7 kategori bu çalışma kapsamında normal ve anormal olmak üzere iki kategoriye indirgenmiştir. Tablo 1' de görüldüğü gibi normal hücrelerin sınıf özelliği 1 olarak, anormal hücrelerin ise 0 olarak etiketlenmiştir. Herlev veri setinde, Şekil 8 (a)' da görüldüğü gibi her sitoplazma bir çekirdek içerdiği için sitoplazmada bulunan çekirdek sayısı ve çekirdek sitoplazma arası mesafe sınıflandırma kriterleri içinde kullanılmamıştır.

Tablo 1'de gösterilen ÇekirdekÇevre ve StpCevre özellikleri çekirdek ve sitoplazma konturlarının çevre hesabını ifade etmektedir. Nrnd ve SRnd özellikleri ise Denklem 7'de verilen denklem ile hesaplanmış olan çekirdek ve sitoplazmaların yuvarlaklık oranını belirtmektedir. Anormal hücrelerin çekirdeklerinde genişleme görüldüğü için ÇekirdekAlan özelliği ile belirtilen çekirdek alanları ile StpAlan özelliği belirtilen sitoplazma alanının birbirine yakın olduğu bu nedenle NC_Oranı olarak belirtilen ÇekirdekAlan/StpAlan oranlarının yüksek çıktığı görülmüştür.

Tablo 1 Normal ve anormal hücelere ait bazı veriler

NC_Oranı	CekirdekAlan	StpAlan	CekirdekÇevre	StpCevre	NRnd	SRnd	Class
1.189425684	4825.5	4057	267.8650047	393.8893892	0.84512359	0.3285	0
0.80497289	3415.5	4243	228.5512966	341.060962	0.8216687	0.4583	0
0.01559438	569	36487	92.76955187	752.055909	0.8906882	0.8106	1
0.15582719	3639.5	23356	599.168608	691.712764	0.6928079	0.6134	1

Oluşturulan veri seti üzerinde makine öğrenmesi yöntemlerinden olan SVM, LR, MLP, RF, NaiveBayes ve KNN yöntemleri uygulanmıştır ve elde edilen sonuçlar Tablo 2’de gösterilmiştir. Değerlendirme işleminde ölçüt olarak doğruluk (accuracy), hassasiyet (precision), duyarlılık (sensitivity) ve özgünlük (specificity) değerleri kullanılmıştır.

SVM yönteminde çekirdek fonksiyonu olarak linear kullanılmıştır ve düzenleme parametresi (C) 1.00 olarak alınmıştır.

MLP yönteminde aktivasyon fonksiyonu olarak ReLU (Rectified Linear Unit) kullanılmıştır ve gizli katman sayısı 3 olarak belirlenmiştir. Her bir gizli katmanda bulunan düğüm sayısı 10 ve 100 değerleri arasında değiştirilerek başarı değerleri ölçülmüştür. Ölçümler sonucunda düğüm sayısının 40 ve üstüne çıktığı sınıflandırma işlemlerinde eğitim setinde başarı artarken test setinde başarı azalmıştır. Bu nedenle her katmanda bulunan düğüm sayısı 30 olarak alınmıştır.

RF yönteminde ormanda bulunan ağaç sayısını belirten *n_estimators* değişkeni 10 olarak alınmıştır.

Tablo 2 Sınıflandırma yöntemleri ve doğruluk değerleri

	SVM	RF	MLP	LR	KNN (k=9)	NaiveBayes
Doğruluk (%)	89.2	92.04	90	87.5	90	83.8
Hassasiyet (%)	82	80	93.1	100	85.2	72.9
Duyarlılık (%)	91.2	95.5	85.7	85.8	90.8	86.3
Özgünlük (%)	82	80	93.1	100	85.2	72.9

Sınıflandırma başarısı olarak RF yöntemi en iyi sonucu vermiştir. Naive Bayes yönteminin başarı oranının düşük olması bu yöntemin istatistiksel veriler ile daha iyi sonuç vermesi olarak düşünülmüştür. MLP yönteminde başarı oranının %90 olarak kalması veri sayısının yeterli miktarda olmaması olarak düşünülmüştür.

Literatürde bu alanda yapılan çalışmalar, kullanılan yöntemler ve elde edilen doğruluk değerleri Tablo 3’te verilmiştir.

Tablo 3 Literatürde Herlev veri seti ile son yıllarda yapılan çalışmalar

Çalışma adı	Sınıflandırıcı	Doğruluk değeri (%)
Chankong vd. [28]	Linear discriminant analysis, KNN, Bayes, SVM, ANN	99.27
William vd. [29]	Fuzzy c-means	98.8
Zhang vd. [30]	ESA	98.3
Win vd.	LD+SVM+KNN+ boosted trees + bagged trees	98.27
Önerilen yöntem	SVM, KNN, RF, LR, MLP, NaiveBayes	92.04

Çalışmanın sınıflandırma başarısı Tablo 3 ile verilen Herlev veri seti üzerinde yapılan diğer çalışmalara göre düşük kalmıştır. Bunun nedeni olarak ise görüntülerden çıkarılan özellik sayısının bu çalışmalarda bulunan özellik sayısına göre az olması düşünülmüştür.

3. Sonuçlar

Bu çalışma kapsamında pap-smear görüntüleri kullanılarak serviks hücrelerinin normal veya anormal olma durumu bulunmuştur. Çalışma segmentasyon, özellik çıkarımı ve sınıflandırma olmak üzere üç ana bölümden oluşmaktadır. Segmentasyon işleminde için renk- tabanlı eşikleme ve K-means olmak üzere iki farklı yöntem kullanılmıştır. Özellik çıkarımında ise görüntülere matematik temelli işlemler uygulanarak sınıflandırma kriterleri belirlenmiştir. Önerilen renk-tabanlı segmentasyon ve matematiksel işlemler ile yapılan özellik çıkarımı yöntemleri doğru çekirdek- sitoplazma eşleşmesi yaptığı için ve hücrelerde bulunan fazla çekirdek sayılarını tespit edip anormal hücreleri görüntü üzerinde bulduğu için birden fazla hücre içeren pap-smear test sonuçlarında patoloji uzmanlarına hastalığın teşhisinde katkı sağlayacaktır.

Oluşturulan veri seti genişletilerek DSIL, YSIL için daha belirleyici ve sayıca fazla kriterler ve daha fazla pap-smear görüntüsü eklenerek rahim ağzı kanserini daha erken ve daha başarılı sonuçlar ile tespit eden bilgisayar-temelli bir arayüz geliştirilerek çalışma tekrar yapılabilir.

Teşekkür

Bu çalışmada rahim ağzı kanseri ve hastalığın teşhisi hakkında bilgi veren İstanbul Zeynep Kâmil Kadın ve Çocuk Hastalıkları Eğitim ve Araştırma Hastanesi Patoloji Bölümü'ne teşekkür ederim.

Referanslar

- [1] "National Institutes of Health homepage", 2020. [Online]. Available: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>. [Accessed: 01-Feb-2020].
- [2] Cancer Today, "Global Cancer Observatory homepage", 2018. [Online]. Available: <http://gco.iarc.fr/today/online-analysis-table>. [Accessed: 02-Feb-2020].
- [3] DB. Cooper, C.E. McCathran, "Cervical Dysplasia", *StatPearls [Internet]*. StatPearls Publishing, 2019.
- [4] RJ. Kurman, D. Solomon, *The Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses*. New York: Springer-Verlag, 1994.
- [5] S.E. Waggoner, "Cervical cancer", *The Lancet*, vol. 361, no. 9376, pp. 2217-2225, 2003.
- [6] T. Bilal, J. Dias, and N. Werghi, "Classification of cervical-cancer using pap-smear images: a convolutional neural network approach", *Annual Conference on Medical Image Understanding and Analysis*, Springer, Cham, 2017.
- [7] M.E. Plissiti, N. Christophoros, and A. Charchanti, "Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering", *IEEE Transactions on information technology in biomedicine*, vol. 15, no.2, pp. 233-241, 2010.

- [8] P. Wang, L. Wang, Y. Li, Q. Song, S. Lv, X. Hu, "Automatic cell nuclei segmentation and classification of cervical Pap smear images", *Biomedical Signal Processing and Control*, vol. 48, pp. 93-103, 2019.
- [9] Y. Marinakis, G. Dounias, and J. Jantzen, "Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification", *Computers in Biology and Medicine*, vol. 39, no.1, pp. 69-78, 2009.
- [10] A. GençTav, S. Aksoy, and S. Önder, "Unsupervised segmentation and classification of cervical cell images", *Pattern recognition*, vol. 45, no.12, pp. 4151-4168, 2012.
- [11] H.A. Phoulady, "A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images", *Computerized Medical Imaging and Graphics*, vol. 59, pp. 38-49, 2017.
- [12] K.P. Win, Y. Kitjaidure, K. Hamamoto, T. Myo Aung, "Computer-Assisted Screening for Cervical Cancer Using Digital Image Processing of Pap Smear Images", *Appl. Sci.*, vol. 10, no.5, pp.1800, 2020.
- [13] W. William, A. Ware, A.H. Basaza-Ejiri, & J. Obungoloch, "A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images", *Biomedical engineering online*, vol.18, no.1, pp.16, 2019.
- [14] V. Vapnik, *The nature of statistical learning theory*. New York: Springer-Verlag, 1995.
- [15] M. Pal, "Random forest classifier for remote sensing classification", *International Journal of Remote Sensing*, vol.26, no.1, pp. 217-222, 2005.
- [16] M.W. Gardner, S.R. Dorling, "Artificial neural networks (the multilayer perceptron) —a review of applications in the atmospheric sciences", *Atmospheric Environment*, vol. 32, no. 14–15, pp. 2627-2636,1998.
- [17] Y. Liao, V. R. Vemuri, "Use of k-nearest neighbor classifier for intrusion detection", *Computers & security*, vol. 21, no.5, pp. 439-448, 2002.
- [18] M. M. Saritas, A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification", *International Journal of Intelligent Systems and Applications in Engineering*, vol. 7, no. 2, pp. 88-91, 2019.
- [19] S.K. Shevade, S. S. Keerthi, "A simple and efficient algorithm for gene selection using sparse logistic regression", *Bioinformatics*, vol. 19, no. 17, pp. 2246-2253, 2003.
- [20] MDE-lab, "MDE-lab downloadspage,"2011.[Online]. Available: <http://mde-lab.aegean.gr/downloads>. [Accessed: 30-Sep-2019].
- [21] J. Jantzen, J. Norup, Jonas, G. Dounias, B. Bjerregaard, "Pap-smear Benchmark Data For Pattern Classification", *Nature Inspired Smart Information Systems (NiSIS)*, pp. 1-9, 2005.
- [22] N. Nill, "A visual model weighted cosine transform for image compression and quality assessment", *IEEE Transactions on communications*, vol.33, no. 6, pp. 551-557, 1985.
- [23] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, A. Charchanti, SIPAKMED: A new dataset for feature and image based classification of normal and pathological

cervical cells in Pap smear images, *IEEE International Conference on Image Processing (ICIP) 2018*, Athens, Greece, 7-10 October 2018.

[24] H. Demirel, G. Anbarjafari, "Discrete Wavelet Transform-Based Satellite Image Resolution Enhancement", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 1997-2004, 2011.

[25] H. Demirel, G. Anbarjafari, "Image resolution enhancement by using discrete and stationary wavelet decomposition", *IEEE transactions on image processing*, vol. 20, no. 5, pp. 1458-1460, 2010.

[26] E.S. Cibas, B.S. Ducatman, *Cytology E-Book: Diagnostic principles and clinical correlates*, Elsevier Health Sciences, 2013.

[27] "Eurocytology Cervical Cytology homepage", 2020. [Online]. Available: <https://www.eurocytology.eu/en/course/1292>. [Accessed: 15-Mar-2020].

[28] T. Chankong, N. Theera-Umpon, & S. Auephanwiriyaikul, "Automatic cervical cell segmentation and classification in Pap smears", *Computer methods and programs in biomedicine*, vol.113, no.2, pp.539-556, 2014.

[29] W. William, A. Ware, A.H. Basaza-Ejiri, & J. Obungoloch, "Cervical cancer classification from Pap-smears using an enhanced fuzzy C-means algorithm", *Informatics in Medicine Unlocked*, vol.14, pp. 23-33, 2019.

[30] L. Zhang, L. Lu, I. Nogues, R.M. Summers, S. Liu, & J. Yao, "DeepPap: deep convolutional networks for cervical cell classification", *IEEE journal of biomedical and health informatics*, vol. 21, no.6, pp. 1633-1643, 2017.