



Mamografi İmgelerimdeki Kitle Tespiti ve Normal-İyi-Kötü Huylu Ayırımı

Mass Detection on Mammograms and Normal-Benign-Malignant Distinction

Nuh ALPASLAN¹, Murat Toptaş¹, Buket Öztürk¹, Davut Hanbay¹

¹Bilgisayar Mühendisliği Bölümü, İnönü Üniversitesi, Malatya
¹e-posta: nuhalpaslan@inonu.edu.tr

Özetçe

Bu çalışmada, mamogramlarda kitlenin belirlenmesi ve bu kitlenin normal, iyi huylu yada kötü huylu olup olmadığının tespiti amaçlanmaktadır. Öncelikle IRMA (Image Retrieval in Medical Applications) veritabanından alınan imgeler üzerinde ön işlem gerçekleştirilmiştir [1]. Uygun bir eşik değeri belirlenerek kitlenin ayrıştırılması sağlanmıştır. Ayrıştırma işleminden sonra gürültülerin elenmesi için morfolojik süzgeçleme yapılmıştır. Bölütleme işleminden sonra tümöre ait geometrik, dokusal ve renksel özellikler elde edilmiştir. Elde edilen bu özellikler YSA ve ELM sınıflandırıcıları kullanılarak, bulunan dokular; normal, iyi huylu ve kötü huylu olmak üzere 3 sınıfa ayrılmıştır. Sonuç olarak bu iki sınıflandırıcının performansları kıyaslanmıştır ve ELM yönteminin YSA'ya oranla daha başarılı sonuçlar verdiği ve daha hızlı çalıştığı gözlemlenmiştir.

Abstract

In this study, it is aimed the determination of the masses in mammograms and the masses whether normal, benign or malignant is intended to determine. Firstly; preprocessing is carried out on IRMA (Image Retrieval in Medical Applications) database images. Mass separation was implemented with an appropriate threshold value. After mass separation, morphological filtering was performed for noise elimination. Then, geometric, textural and color features of tumors have been obtained. The obtained features are classified into 3 classes as normal, benign or malignant using ANN and ELM classifiers. As a result performance of these two classifiers have been compared and it is seen that, ELM method is more succesful and faster than ANN.

1. Giriş

Meme kanseri, tüm dünyada, kadınlarda sıkça, erkeklerde ise seyrek olarak görülen ve ölüme yol açan bir kanser türüdür. Ülkemizde ve dünyada kadınlar üzerinde en sık görülen bu kanser, ortalama olarak 45 yaş ve üstünde kendini göstermektedir. Türkiye'de meme kanserinin 1993 yılındaki ortalama insidansı yüz binde 24 iken, 2010 yılında bu oran yüz binde 50 civarındadır [2].

Maliyetinin düşük olması ve erken teşhiste etkili olmasından ve 40 yaşını geçmiş kadınlarda yılda en az bir kez mamografi çektirmenin doktorlarca tavsiye edilmesinden dolayı, mamografiye başvuru artmış ve bu durum bir doktorun incelemesi gereken vaka sayısının da doğru orantılı olarak artmasına neden olmuştur.

2006 yılında Avrupa'da yaklaşık 3.191 milyon kanser vakası teşhis edilmiştir. En fazla görülen kanser, %13.5 (429900 vaka) ile göğüs kanseridir. Her yıl Avrupa'da 3.2 milyon yeni kanser vakası oluşacağı tahmin edilmektedir [3]. Bu kapsamda elde edilen verilerin otomatik olarak analiz edilmesi ve sınıflandırılması hem hastalar hem de sağlık sektörü açısından büyük önem taşımaktadır.

Mamogramlardan daha iyi yararlanabilmek ve teşhislerdeki doğruluk oranını arttırabilmek için, doktora yardımcı olacak bilgisayar destekli bir karar destek sisteminin geliştirilmesi, hem süreci kısaltacak hem de tanının doğru bir şekilde verilmesine yardım edecektir.

Bu çalışmada mamogramlarda tümörün tespit edilmesinde ilk olarak bölütleme işlemi sonucunda arama yapılacak alan elde edilmektedir. Sonrasında mamogramlarda meme bölgelerinin elde edilmesi için yoğunluk ve renk özelliklerinden yararlanılmaktadır. Yoğunluk ve renk değerleri belirlenen dinamik aralığa göre sınırlandırılmaktadır. Dinamik aralık sonucunda oluşan ikili imgede matematiksel morfoloji adımları uygulanıp istenilen alan elde edilmektedir. İkinci aşamada önceki aşamada elde edilen alan üzerinden geometrik, dokusal ve renksel özellikler elde edilmiştir. Üçüncü ve son aşamada ise bulunan özellikler yardımıyla vakalar sağlıklı, iyi huylu kanser, kötü huylu kanser olmak üzere üç sınıfa ayrılmıştır.

2. Yöntem

2.1. Ön İşlem Aşaması

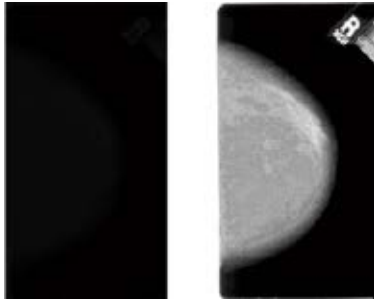
2.1.1. Histogram Germe

Kontrast bozukluğu olan resimlerin histogram dağılımı belirli bir aralıkta yoğunlaşmıştır. Renk seviyeleri birbirine yakın olduğundan görüntünün algılanması da zor olmaktadır. Histogram germe işlemi gerçekleştirilmek için her noktaya ait renk değerini aşağıdaki fonksiyona tabii tutmak gerekir.

$$Y_i = (E_i - E_{min}) * (Y_{max} - Y_{min}) / (E_{max} - E_{min}) + Y_{min}$$

Tıbbi Görüntüleme 2

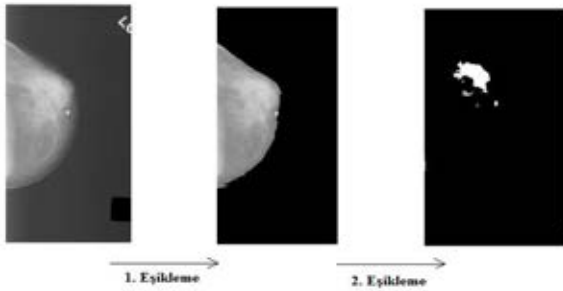
3. Gün 27 Eylül 2014 Cumartesi (11.45-13.00)



Şekil 1- Histogram Germe İşlemi

2.1.2. Görüntü Bölütleme

Eşikleme, arkaplan ile nesneyi ayırt etmede kullanılan bir yöntemdir. Bu çalışmada iki adımlı eşikleme kullanılmıştır. Eşikleme seviyeleri OTSU eşikleme seviyesi algoritmasıyla hesaplanmıştır [4]. Birinci eşikleme ile göğsün arkaplandan ayırt edilmesi amaçlanmıştır. İkinci eşikleme ile ise bulunan göğüs imgesi içerisindeki kitlenin tespiti amaçlanmıştır. Böylece mamogram çekilirken arkaplanda meydana gelen gürültüler, kitlenin tespitinde etkili olmadığından daha başarılı sonuçlar elde edilmiştir.



Şekil 2- Görüntü Bölütleme

2.1.3. Gürültülerin Elenmesi ve Şüpheli Doku Tespiti

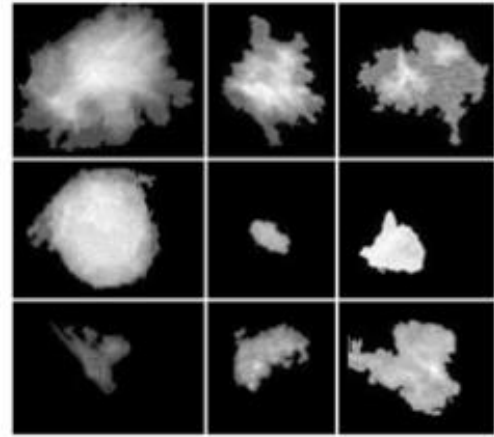
Eşiklemeler sonunda elde ettiğimiz alan etrafında çeşitli gürültüler de birlikte gelebilmektedir. Bu gürültüler morfolojik süzgeçleme kullanılarak elenerek şüpheli kitle tespit edilir.



Şekil 3- Gürültülerin Elenmesi

2.2. Öz nitelik Çıkarımı

Nesne sınıflandırmada öz nitelik çıkarımı kritik bir rol oynamaktadır. Çıkarılan öz niteliğin sınıflandırılacak olan nesneyi ayırt edici kılması gerekmektedir.



Şekil 4- a) kötü huylu doku, b) iyi huylu doku, c) normal doku

Öz nitelik çıkarma aşamasından önceki aşamalar (görüntü iyileştirme, görüntü bölütleme, doku tespiti) sonucunda elde edilen doku örnekleri şekil 1'deki gibidir. Özellikler genel olarak geometrik, dokusal ve renksel olarak 3 grupta incelenebilir.

2.2.1. Geometrik Özellikler

Şekil 1'deki doku örneklerinde gözüktüğü üzere şüpheli kitlenin iyi huylu, kötü huylu veya sağlıklı olması kenarlarıyla doğrudan ilişkilidir. Geometrik özelliklerin tespit edilebilmesi için öncelikle kenarların tespit edilmesi gerekir. Bu çalışmada şüpheli dokuya ait kenarlar sobel kenar bulma yöntemi ile tespit edilmiştir [5]. Kenarların ne kadar düzenli dağılımda bulunduğunu tespit etmek üzere 3 geometrik özellikten yararlanılmıştır.

2.2.1.1 Kenarların merkeze olan uzaklıkları varyansı

Varyans bir olasılık dağılımı veya örneklem için istatistiksel yayılımın, mümkün bütün değerlerin beklenen değer veya ortalamadan uzaklıklarının karelerinin ortalaması şeklinde bulunan bir ölçüdür. Kitle kenarlarının merkezden uzaklığının dağılımını incelemek bize kitlenin yuvarlaklığı hakkında bilgi verecektir. Bu çalışmada Kenarların merkezden uzaklığı öklit uzaklığı ile hesaplanmıştır.

$$d_{|PM|} = \sqrt{(Px - Mx)^2 + (Py - My)^2}$$

Bulunan değerlerin diğer dokularla karşılaştırıldığında büyüklük farklarından etkilenmemesi için bu değerler normalize edilir.

$$dn_k = \frac{(d_k - d_{min})}{d_{max} - d_{min}}$$

Normalize edilmiş uzaklıkların ortalaması dn_{ort} 'dan yararlanılarak kenarların merkeze olan uzaklıklarının varyans değeri tespit edilir.

$$Varyans = \frac{\sum_{i=0}^N (dn_i - dn_{ort})^2}{N - 1}$$



Tıbbi Görüntüleme 2

Varyans değeri, kitlenin kenarlarının (merkezden uzaklığının) ne kadar düzensiz dağılıma sahip olduğu hakkında bilgi verir.

2.2.1.2 Alanın Yuvarlaklık Oranı

Her kenar noktasının merkeze olan uzaklığı, ortalama uzaklıktan çıkartılarak toplanır. Böylece merkezi kitle merkezi olan, ortalama uzaklık yarıçaplı bir dairenin alanından farkı bulunur. Bulunan fark alanı, bu çemberin alanına bölünerek farkın oranı tespit edilir.

$$f(x) = \frac{1}{dn_{ort} \cdot N} \sum_{i=1}^N (dn_i - dn_{ort})$$

2.2.1.3 Pürüzsüzlük

Kenar üzerindeki komşu iki nokta arasındaki uzaklık farklarının ortalamasıdır. Bu özellik merkeze olan uzaklık değerlerinin en küçük aralıktaki değişimlerini temsil eder. Noktaların merkeze olan uzaklıklarının ani değişimini tespit etmede faydalıdır.

$$(x) = \frac{1}{N} \sum_{i=0}^{N-1} |dn_i - dn_{i+1}|$$

2.2.2. Dokusal Özellikler

Dokusal özellikleri elde etmek için öncelikle, imgenin GLCM eş oluşum matrisi elde edilir. GLCM iki komşu piksel arasındaki ilişkiyi ifade etmekte ve bunların tekrar sayılarını içermektedir. Dokusal özellikler elde edilen yeni matris üzerinden hesaplanır.

2.2.2.1 Normalize Moment Farkının Tersisi

Tekrar eden değerlerin arasındaki farkın yüksek olma durumunun formül üzerindeki etkisi daha küçük olacağından, bu değer ne kadar büyükse tekrar eden örüntü içerisinde birinci değer ile ikinci değer o kadar birbirine yakındır anlamına gelmektedir. Normalize moment farkı tersi bize sıralı olarak tekrar eden değerlerin birbirine yakınlığı hakkında, yani piksel renk geçişlerinin sert veya yumuşak olduğu hakkında bilgi vermektedir.

$$f(x) = \sum_{i=1}^N \sum_{j=1}^N \frac{1}{1 + (i - j)^2} P(i, j)$$

2.2.2.2 Entropi

Entropi düzensizlik olmakla birlikte dağılımı fazla olan noktalar için daha büyük, dağılımı az olan noktalar için daha küçük sonuçlar vermektedir. Sonuç olarak az olan değeri daha az etkileyerek, çok olan değeri daha çok etkileyerek aralarındaki farkı iyice artırır. Bu özellik ile örüntülerin ne kadar çok tekrar edip etmediği hakkında bilgi sahibi oluruz.

$$f(x) = - \sum_{i=1}^N \sum_{j=1}^N P(i, j) \log P(i, j)$$

2.2.2.3 Enerji

Enerji graycomatristeki tüm değerlerin karelerinin toplamına eşittir. Bütün değerler birbiriyle aynı olduğu durumda minimum olur. Küçük bir değere eşit olması dağılımın daha düzgün olduğunu gösterir. Enerji değerinin büyümesi, farklı

3. Gün 27 Eylül 2014 Cumartesi (11.45-13.00)

örüntülerin aynı sayıda tekrar etmiş olmaktan uzaklaştığı anlamına gelmektedir.

$$f(x) = \sum_{j=0}^N \sum_{i=0}^N P^2(i, j)$$

2.2.2.4 Korelasyon

Değerlerin ve bu değerlerin ardı ardına gelme sayısının çarpılıp ortalamanın karesinden farkı bulunur, daha sonra bu değer varyansa bölünerek korelasyon hesaplanır. Değerlerimizin veya bu değerlerin ardı ardına tekrar etme sayılarının fazla olması değerimizi artıracaktır. Düzenli bir dağılıma sahip olması, eş sayıda birbirini takip eden değerlerin bulunması varyans değerini düşüreceğinden değerlerin büyümesine sebep olacaktır.

$$f(x) = \frac{\sum_{i=1}^N \sum_{j=1}^N (ij)P(i, j) - Ort^2}{Varyans}$$

2.2.3. Renkisel Özellikler

2.2.3.1 Gri Eşikleme Seviyesi

Mamogram görüntüsü göğüs dokusunun yağ-bağ doku oranına bağlı olarak daha parlak veya daha mat olabilmektedir. Parlak görüntülerde kitlenin ve kanserin teşhisi daha zor olmakla birlikte, göğüsün yapısının (doku oranının) kanser üzerinde etkili olduğu düşünülmektedir. Bu sebeple tüm göğüze ait gri eşikleme seviyesi özellik olarak kullanılmaktadır.

2.3. Sınıflandırma

Bu çalışmada sınıflandırıcı olarak yapay sinir ağları (YSA) ve aşırı öğrenme makinaları (ELM) yöntemleri kullanılmış ve yöntemlerin sınıflandırma performansları değerlendirilmiştir. YSA, insan beyninden esinlenerek, öğrenme sürecinin matematiksel olarak modellenmesini herhangi bir yardım almadan otomatik olarak gerçekleştirebilmek amacı ile geliştirilen bilgisayar sistemleridir. Aşırı öğrenme makinası giriş ağırlıklarının rastgele, çıkış ağırlıklarının analitik olarak hesaplandığı tek gizli katmanlı ileri beslemeli özel bir YSA modelidir.

Geleneksel ileri beslemeli yapay sinir ağlarında ağırlıklar ve eşik değerleri gibi parametrelerin gradyan tabanlı öğrenme algoritmaları ile güncellenmesi gerekmektedir. Ancak iyi performansın sağlanması için öğrenme süreci hem zaman almakta hem de hata yerel bir noktaya takılabilmektedir. Bunun için momentum değerinin değiştirilmesi belki hatanın lokal bir noktaya takılmasını engelleyebilir ancak öğrenme sürecinin süresini etkilemeyecektir. Tek gizli katmanlı ileri beslemeli bir ağda giriş-çıkış ağırlıkları ve eşik değerleri ağırlık performansını etkilememektedir. ELM de giriş ağırlıkları ve eşik değerleri rastgele üretilmekte ancak çıkış ağırlıkları hesaplanmaktadır. Bu yüzden ELM geleneksel yöntemlere göre daha performanslı ve hızlı çalışmaktadır.

3. Deneysel Sonuçlar

Bu çalışmada IRMA (Image Retrieval in Medical Applications) projesi kapsamında hazırlanan Digital Database for Screening Mammography (DDSM) veri seti kullanılmıştır [1]. Veri setinde, her hastaya ait sağ ve sol olmak üzere 2 göğsün, 2'şer farklı açıdan çekilmiş toplamda 4 adet mamogram görüntüleri bulunmaktadır. Toplam hasta sayısı 2446'dır. Bunlardan 913'ü

Tıbbi Görüntüleme 2

3. Gün 27 Eylül 2014 Cumartesi (11.45-13.00)

kötü huylu (kanser), 821'i iyi huylu ve 712'si sağlıklıdır. Bu verisetinin %85 lik kısmı eğitim, kalan %15 lik kısmı ise test işleminde kullanılmıştır.

Bu çalışmada bahsedilen tüm algoritmalar 2,5-GHz Intel(R) i5 CPU, 4GB Ram konfigürasyonlu bir bilgisayar üzerinde Matlab programı kullanılarak uygulanmıştır.

Kullanılan verisetinde her bir hastaya ait toplamda 4 adet görüntü bulunmaktadır. Bir hastaya ait olan 4 görüntü için elde edilen özelliklerin ayrı ayrı kullanılması ile 32 girişli sınıflandırıcı modeli, elde edilen özelliklerin sağ ve sol göğsün kendi içlerinde toplanarak oluşturulan 16 girişli sınıflandırıcı modeli ve son olarak; 4 resme de ait özelliklerin toplanarak elde edilen 8 girişli sınıflandırıcı modeli olmak üzere toplamda 3 adet sınıflandırıcı modeli denenmiştir. Elde edilen doğruluk ve çalışma süresi tabloları Tablo 1 ve Tablo 2 'de verilmiştir.

Tablo 1 de YSA ve ELM'ye ait sınıflandırma başarıları görülmektedir. Tablo 1 incelendiğinde ileri beslemeli YSA sınıflandırıcısı için en yüksek %61 ELM için ise en yüksek %71 başarı oranı elde edildiği görülmektedir. Ayrıca tablo 1 incelendiğinde kullanılan özellik vektörü sayısı arttıkça sınıflandırma başarısının arttığı görülmektedir.

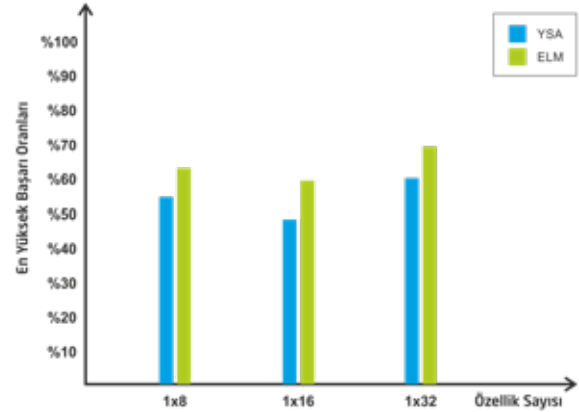
Tablo 2 ise YSA ve ELM sınıflandırıcılarına ait ortalama çalışma sürelerini ifade etmektedir. Tablo 2 incelendiğinde ELM yönteminin YSA yöntemine oranla yaklaşık olarak 13 kat daha hızlı çalıştığı görülmektedir.

	YSA	ELM
1x8	%56	%61
1x16	%50	%60
1x32	%61	%71

Tablo 1- Sınıflandırıcılara ait doğruluk değerleri (Accuracy)

	YSA	ELM
1x8	0.850	0.090
1x16	1.350	0.080
1x32	0.950	0.070

Tablo 2- Sınıflandırıcıların ortalama çalışma süreleri (sn)



Şekil 5- Kullanılan sınıflandırıcıların başarı oranları

4. Sonuçlar

Bu çalışmada, mamogram imgelerindeki kitlenin belirlenmesi ve bu kitlenin normal, iyi huylu yada kötü huylu olup olmadığını tespiti ve sınıflandırılması gerçekleştirilmiştir. Bu amaçla ilk olarak IRMA (Image Retrieval in Medical Applications) veritabanından alınan imgeler üzerinde ön işlem gerçekleştirilmiştir. Daha sonra uygun bir eşik değeri belirlenerek imge üzerindeki kitle ayrıştırılarak istenilen bölgenin elde edilmesi sağlanmıştır. İstenilen kitle ayrıştırıldıktan sonra gürültülerin elenmesi için morfolojik süzgeçleme yapılmıştır. Bu aşamadan sonra her bir kitleye ait 3 adet geometrik, 4 adet dokusal ve 1 adet renksel özellikler elde edilmiştir. Elde edilen bu özellikler YSA ve ELM sınıflandırıcıları kullanılarak, normal, iyi huylu ve kötü huylu olmak üzere 3 sınıfa ayrılmıştır. Sonuç olarak sınıflandırıcıların performansları incelendiğinde ELM yönteminin YSA'ya göre daha yüksek sınıflandırma başarısına sahip olduğu görülmüştür. Ayrıca, sınıflandırıcıların çalışma süreleri incelendiğinde ELM yönteminin YSA'ya göre yaklaşık olarak 13 kat daha hızlı çalıştığı gözlemlenmiştir.

5. Kaynakça

- [1] Lehman TM, Oliveira JEE, Güld MO, Weter P., "Data Digital Database For Screening Mammography (DDSM) in PNG Format", "IRMA Version Of DDSM LJPEG"
- [2] Prof. Dr. Vahit Özmen, Dünyada ve Türkiye'de Meme Kanseri, İstanbul Üniversitesi Tıp Fakültesi, 20. Ulusal Kanser Kongresi, 2013, Antalya
- [3] Ferlay, J., Autier, P., Boniol, M., Heanue, M., Colombet M., Boyle P., "Estimates of the cancer incidence and mortality in Europe in 2006", Annals of Oncology, 18 581-592, 2007.
- [4] Bryn S. Morse, Lecture 4: Thresholding, Brigham Young University, 1998-2000.
- [5] Elif Aybar, Sobel Edge Detection Method For MATLAB, Anadolu Üniversitesi, Eskişehir,