



PROSTAT HÜCRE ÇEKİRDEKLERİNİN SINIFLANDIRILMASINDA İSTATİSTİKSEL YÖNTEMLERİN VE YAPAY SINIR AĞLARININ BAŞARIMI

*(PERFORMANCE OF STATISTICAL METHODS AND ARTIFICIAL
NEURAL NETWORKS FOR CLASSIFICATION OF PROSTATE CELL
NUCLEI)*

Metehan MAKİNACI*

ÖZET/ABSTRACT

Bu çalışmada, patolojik verilere uygulanan farklı sınıflandırıcıların başarımları analiz edilmiştir. 80 normal ve 80 kanserli prostat hücre çekirdek imgesinden, Gauss Markov rassal alanlar, Fourier entropi ve dalgacık dönüşümü ortalama sapma öznelik vektörleri elde edilmiş ve ayırım gücü olanlardan ortak bir öznelik vektörü oluşturulmuştur. Sınıflandırma için yapay sinir ağları, k-en yakın komşu, en yakın merkez ve doğrusal ayırtaç yöntemleri kullanılmıştır. Bu aşamada, 3-15 arası ara katman düğümüne sahip geri yayımlı yapay sinir ağı, sınıflandırma amacı ile eğitilip test edilmiştir. En yüksek genel başarımlar oranını %85.5 ile en yakın merkez sınıflandırıcısı sağlamıştır.

In this study, performances of different classifiers were analyzed for pathological data. Gauss Markov random field, Fourier entropy, and wavelet mean deviation features were calculated for 80 normal and 80 cancerous prostate cell nuclei and a common feature set was created from the ones having the discrimination power. Neural networks, K-nearest neighbor, nearest mean, and linear discriminant classifiers were used for classification. In this stage backpropagation neural networks having 3 to 15 hidden layer nodes were trained and tested. Highest classification rate (85.5%) was achieved by the nearest mean classifier.

ANAHTAR KELİMELELER/KEYWORDS

Doku sınıflandırma, Dalgacık dönüşümü, Markov rassal alanlar, Yapay sinir ağları
Texture classification, Wavelet transform, Markov random fields, Artificial neural networks

1. GİRİŞ

Kanser çağımızın önemli bir hastalığıdır. Bu hastalığın tedavi edilebilmesi için ilk aşamada doğru tanı koymak gerekmektedir. Kanser tanısının kesin olarak konabilmesi ancak alınan örneklerin mikroskop altında bir patolog tarafından incelenmesi ile mümkün olmaktadır. Patolog, inceleme sonucu elde ettiği bilgileri, koyduğu tanı ile birlikte tedaviden sorumlu doktora iletir. Doktor, bu bilgileri analiz ederek en uygun tedavi yöntemine karar verir. Başarılı bir tedavi için doğru tanının önemi çok büyüktür. Bazı vakalarda kesin tanı koymak oldukça zor bir problemdir ve uzman görüşleri arasında %30-40 farklılık olabilmektedir. Bu tip durumlarda nesnel ölçütleri kullanan yardımcı yöntemler, tanının başarımını arttırmada uzmana yardımcı olacaktır.

Bir sınıflandırma problemine çözüm bulmak için en iyi nesnel ölçütlerin belirlenmesi gerekmektedir. Amaç, eldeki veriyi en iyi şekilde matematiksel olarak modellemektir. Doku modelleme yöntemlerinden bazıları, Fourier güç spektrumu, birinci ve ikinci derece gri seviye istatistikleridir (Weszka vd., 1976, Haralick 1979). Bu çalışmalardan yola çıkılarak geliştirilen fraktal boyut, Gabor süzgeçleri, ve dört evreli süzgeçler (Du Buf vd., 1990) hep insanın görme mekanizmasını modellemeyi amaçlamıştır. Yapılan araştırmalar sonucunda, insan görme sisteminin, belli yönelime ve uzamsal frekansa duyarlı, birbirinden bağımsız birçok kanal gibi modellenebileceği görülmüştür. Doku görüntüsünün insan görme sistemi tarafından çok ölçekli olarak işlenmesi, benzer özelliklere sahip matematiksel modellerin doku analizinde kullanılmasının temel sebebini oluşturmaktadır (Bovik, vd., 1990, Van de Wouwer vd, 1999). Dalgacık dönüşümü de bu özelliklere sahip bir modelleme yöntemidir ve bu çalışmada Fourier entropi modeli ile birlikte kullanılmıştır.

Modellemede kullanılan üçüncü yöntem Gauss Markov rassal alanlarıdır (GMRA). İki boyutlu zaman serileri ve rassal alanlar en sık kullanılan imge modelleri arasındadır. Bu grupta yer alan oto model, çok-düzeyleli lojistik model ve sıradüzensel Gibbs modeli (Li,1995) imgedeki yerel istatistiksel bilgiyi karakterize eder. Bu açıdan bakılarak, patolojik imgelerin 2 boyutlu rassal alan ile modellenmesi sonucunda istatistiksel öznitelikler elde edilmiştir.

Problemin çözümü için, bir sonraki aşamada, en iyi sınıflandırma yönteminin seçilmesi gerekmektedir. Sınıflandırma yöntemleri gözetici ve gözetici olmayan üzere ikiye ayrılır. Gruplama, istatistiksel olan ve istatistiksel olmayan sınıflandırıcılar kullanılarak da yapılır. Bu çalışmada tümü de gözetici sınıflandırma yöntemlerinden olan üç istatistiksel sınıflandırıcı ve bir yapay sinir ağı sınıflandırıcı normal ve kanserli prostat hücre çekirdeklerini sınıflandırmak için kullanılmış ve başarımları analiz edilmiştir.

2. MATERYAL ve YÖNTEM

Bir örüntü tanıma sistemi temelde üç parçadan oluşur. Bunlar: (a) veri alma ve ön işleme, (b) veriyi temsil edecek öznitelikleri çıkarma ve (c) sınıflandırmadır. Bu çalışmada kullanılan sistem bileşenleri aşağıda açıklanmaktadır.

2.1. Histoloji

Bu çalışmada, hastalardan alınan prostat parçalarından elde edilmiş kesitler kullanılmıştır. Patoloji laboratuvarında uzmanlar tarafından gerçekleştirilen hazırlık işlemleri şu şekilde özetlenebilir: Parçalar önce %10'luk formalin çözeltisinde 24-48 saat bekletilerek sabitlenmiştir. Rutin biyolojik doku incelemesinden sonra elde edilen parafin bloklar, 5-µm kalınlığında kesilerek hematoksilin-eosin ile boyanmıştır.

2.2. İmge Kayıt Sistemi

Prostat doku kesitlerinin görüntüsü Nikon Labophot-2 mikroskop kullanılarak x100 büyütülmüştür. Mikroskoba bağlı renkli kamera (Sony DXC-107AP) ile elde edilen analog görüntüler, bilgisayarda bulunan sayısallaştırıcı kart ve yazılım aracılığıyla 768x576 piksel boyutunda 16 milyon renk olarak kayıt edilmiştir.

Patoloji uzmanının etiketlediği 80 normal ve 80 kanserli hücre çekirdek görüntüsünden 32x32 piksel boyutunda bloklar alınmıştır. Seçilen çekirdeklerin net olmasına, çok koyu renk olmamasına, görsel bozukluklar içermemesine, ayrık olmasına ve bloğu tamamen kapsamasına dikkat edilmiştir. Seçilen çekirdek imgeleri gri tona çevrilmiş ve öznelik çıkarma işlemi için kayıt edilmiştir.

2.3. Öznelik Vektörü Hesaplama

Prostat hücre çekirdek imgeleri üç yöntem ile modellenmiştir. Bu modellerden ilki Gauss Markov rassal alanlar modelidir (Chatterjee, 1991, Manjunath, vd., 1991, Li, 1995). İkinci model bağıl entropi (Jernigan, vd.,1984) yöntemi ile oluşturulmuştur. Bu modelde entropi parametreleri, ayrık Fourier dönüşümü ile elde edilen normalize enerji izgesinden elde edilmiştir. Son yöntemde, durağan dalgacık dönüşümünden elde edilen ortalama sapma öznelikleri (Van de Wouwer vd., 1999) hesaplanmıştır. Durağan dalgacık dönüşümünde seyrek örnekleme yapılmadan filtreler uygulanmaktadır. Bu aşamada, ayırıştırma ve yeniden oluşturma için farklı süzgeçler kullanan çift dikgen dalgacıklar (Misiti vd, 1996, Aldroubi, 1996) dönüşüm amacı ile kullanılmıştır.

2.4. Öznelik Seçimi

Ortak öznelik vektörü oluşturmak için, toplam 28 öznelikten (GMRF 12, dalgacık dönüşümü ortalama sapma 9, Fourier entropi 7) en iyi olanlar ($p < 0.001$), t-test ile seçilmiştir (p-seviyesi, gözlenen sonucun geçerliliğindeki hata olasılığını vermektedir.) Bu vektör 7 GMRA, 1 dalgacık dönüşümü ortalama sapma ve 2 Fourier entropi özneliğinden oluşmaktadır. Bu vektör içinden en iyi ayırım gücüne sahip öznelik kombinasyonunu bulmak ve vektör boyutunu azaltmak için ardışıl ileri seçim yöntemi (Jain, vd., 2000) kullanılmıştır. Ardışıl ileri seçim yönteminde öncelikle en iyi öznelik seçilmiş, daha sonra, her defasında bir öznelik eklenerek kriter fonksiyonunu eniyileştiren öznelik bileşimi bulunmuştur. Elde edilen vektör 3 bileşene sahiptir ve her modelden birer öznelik içermektedir.

2.5. Sınıflandırma Yöntemleri

Sınıflandırma, bir örüntü tanıma sisteminin son aşamasında bulunmaktadır ve bu çalışmada prostat hücre çekirdeklerine tanı (kanser/normal) koymak amacı ile kullanılmıştır. Sınıflandırıcıların tümü de gözetilidir ve özellikleri Çizelge 1'de özetlenmiştir. Doğrusal ayırtaç analizi ile sınıflandırıcı oluşturulurken, gruplar arası varyansın en-çoklanması, grup içi varyans ortalamasının da en-azlanması gerekmektedir. Bunun için bir optimizasyon yapılmaktadır. Sınıflandırıcıların, kullanılmadan önce eğitilmeleri gerekmektedir. Eğitim sonucunda sınıflandırıcı, öznelik uzayındaki örüntü sınıfları arasında karar sınırları oluşturur. Sınıflandırıcı istatistiksel ise, sınırlar oluşturulurken istatistiksel karar teorisi (Schürmann,1996) kullanılır. Bayes karar kuralı, en büyük olabilirlik kuralı ve Neyman-Pearson kuralı (Jain, vd., 2000, Duda, vd. 1973) iyi bilinen karar kurallarıdır.

Çizelge 1. Kullanılan sınıflandırıcıların özellikleri (OKH: ortalama karesel hata)

Sınıflandırıcı	Özellik	İstatistiksel mi?
Yapay Sinir Ağı	İki veya daha fazla katmanlı sinir ağının sigmoid transfer fonksiyonlarını kullanarak iteratif OKH optimizasyonu.	Hayır
Doğrusal Ayırtaç	OKH optimizasyonu kullanan doğrusal sınıflandırıcı.	Evet
K-En Yakın Komşu	Başarımı optimize edilmiş bir k değerini kullanarak örüntüyü komşular arasında çoğunluğa sahip gruba atar.	Evet
En Yakın Merkez	Örüntüyü, merkezi en yakın gruba atar.	Evet

2.6. Sınıflandırıcı Başarımı

Sınıflandırıcıların başarımı birini-dışarıda-bırak yöntemi ile ölçülmüştür. Bu yöntemde, bir çekirdek öznitelik vektörü test amacı ile ayrılır ve sınıflandırıcı geri kalanların tümü kullanılarak eğitilir. Eğitim sonucunda, ayrılan vektör ile sınıflandırıcı test edilir ve başarımı ölçülür. Genel başarımı hesaplamak için bu işlem her öznitelik vektörü için uygulanır. Sınıflandırıcı başarımı üç temel bileşene sahiptir: Hassasiyet (kanseri çekirdeği kanserli olarak sınıflandırma oranı), belirleyicilik (normal çekirdeği normal olarak sınıflandırma oranı) ve genel başarım (kanseri ve normal çekirdekleri doğru olarak sınıflandırma oranı). Bu çalışmada, sınıflandırıcı başarımlarını karşılaştırmak için, hassasiyet, belirleyicilik ve genel başarım oranları yanında sınıflandırıcı etkinlik eğrileri (Kil, vd., 1996) de kullanılmıştır. Bu eğriler, hassasiyet ve belirleyicilik oranları kullanılarak oluşturulmuştur. Doğru sınıflandırma ile yanlış sınıflandırma başarımları arasındaki ilişkiyi grafiksel olarak açıklayan eğriler sonuçlar bölümünde verilmiştir.

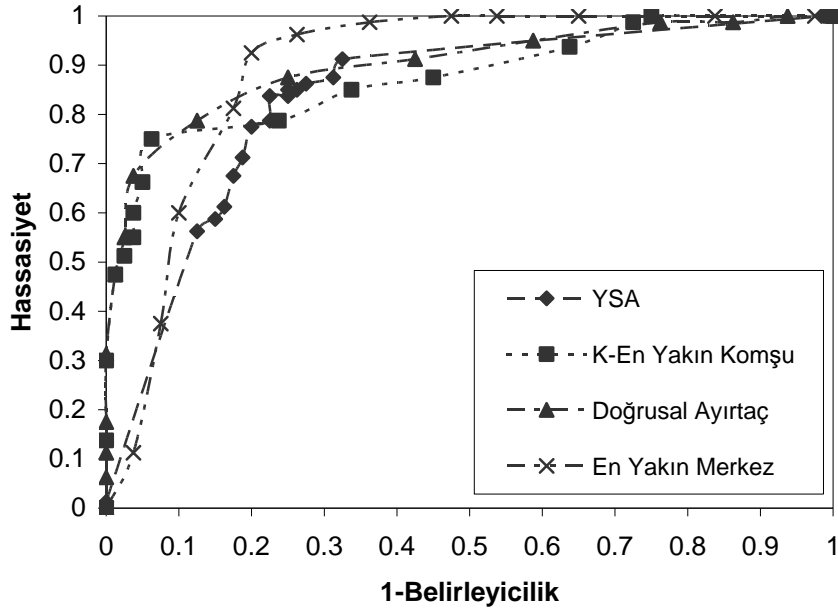
3. SONUÇLAR ve TARTIŞMA

Bu çalışmada, prostat hücre çekirdeklerini normal/kanser olarak ayırmak için kullanılan sınıflandırıcıların başarımları ölçülmüştür. Sınıflandırmada kullanılan öznitelik vektörü, GMRA, dalgalık dönüşümü ve Fourier entropi modellerinden elde edilen birer öznitelikten oluşmaktadır. Sınıflandırıcının 0.5 eşik değeri için, en iyi genel başarım oranını (%84.38) k-en yakın komşu sınıflandırıcısı k=18 ile sağlamıştır. En yakın merkez, doğrusal sınıflandırıcı ve yapay sinir ağı sınıflandırıcılarının genel başarım oranları sırası ile %83.75, %82.50 ve %80.63 olarak hesaplanmıştır. Burada kıstas olarak kullanılabilir olan 1-en yakın komşu sınıflandırıcı %78.75 genel başarım oranı vermiştir. Sınıflandırıcıların hassasiyet, belirleyicilik ve genel başarım oranları Çizelge 2’de özetlenmiştir. Geri yayımlı yapay sinir ağı sınıflandırıcının en iyi sonuç verdiği yapı tek ara katmandan oluşmakta ve bu ara katmanda 13 düğüm elemanı bulunmaktadır. Aynı genel başarım oranını veren diğer bir yapay sinir ağı sınıflandırıcı da 4 ara katman düğümüne sahiptir. Fakat bu sınıflandırıcının hassasiyet ve belirleyicilik oranları %85 ve %76.25 olarak hesaplanmıştır.

Başarımları karşılaştırmak için oluşturulan sınıflandırıcı etkinlik eğrileri Şekil 1’de verilmiştir. Eğrilerin altında kalan alan, sınıflandırıcı başarımı için önemli bir göstergedir. Alan değeri 1 ise, bu tam (%100) başarımı, 0.5 ise başarısızlığı (%50) gösterir. Bu çalışmada kullanılan sınıflandırıcıların alan değerleri, YSA için 0.8354, k-en yakın komşu için 0.8768, doğrusal ayırtaç için 0.8999 ve en yakın merkez için 0.8900 olarak hesaplanmıştır. Bu açıdan bakıldığında doğrusal ayırtaçın en iyi başarıma sahip sınıflandırıcı olduğu görülmektedir.

Çizelge 2. Kullanılan sınıflandırıcıların başarımları (eşik=0.5).

Sınıflandırıcı	Hassasiyet (%)	Belirleyicilik (%)	Genel (%)
Yapay Sinir Ağı	86.25	75.00	80.63
Doğrusal Ayırtaç	78.80	86.30	82.50
K-En Yakın Komşu	75.00	93.75	84.38
En Yakın Merkez	85.00	82.50	83.75



Şekil 1. Sınıflandırıcı etkinlik eğrileri.

Genel olarak, elde edilen sınıflandırıcı başarımları arasında çok büyük farklılıklar bulunmamaktadır. Bunun temel nedeni, kanserli ve normal bir kısım çekirdek verisinin öznitelik uzayında aynı bölge içinde bulunmasıdır. Bu bölgedeki çekirdekler kanserli olarak sınıflandırılacak şekilde sınıflandırıcı eşik değeri belirlenirse hassasiyet oranı artar, fakat aynı zamanda belirleyicilik oranı azalır (Şekil 1'de sağ üst bölgeye kayılır). Aynı çekirdekler normal olarak sınıflandırılacak şekilde eşik değeri ayarlanırsa, hassasiyet ve belirleyicilik oranlarında bir önceki durumun tam tersi bir değişim gözlenir. Hassasiyetin düşük olması kanserli vakaya yanlış tanı koyma olasılığının yüksek olması anlamına gelir. Sabit bir hassasiyet değeri için başarımlara bakılacak olursa, %80 için en iyi genel başarıma %83.5 ile doğrusal ayırtaç sınıflandırıcı sahip olmuştur. Hassasiyetin %90 olması durumunda, en iyi genel başarımları %85.5 ile en yakın merkez sınıflandırıcısı göstermektedir. Bu iki sonuç da eğri alan değerleri ile uyumaktadır. Bu çalışmada elde edilen sonuçlar, uygulanan örüntü tanıma yönteminin patoloji alanında uzmana yardımcı bir sistem olarak kullanılabileceğini göstermektedir.

TEŞEKKÜR

Patolojik görüntülerin sağlanmasında ve etiketlenmesinde yaptığı katkılardan dolayı D.E.Ü., Tıp Fakültesi, Patoloji A.B.D.'dan Prof. Dr. Kutsal Yörükoğlu'na teşekkür ederim. Bu çalışma, D.E.Ü. 02.KB.FEN.058 nolu proje kapsamında gerçekleştirilmiştir.

KAYNAKLAR

- Aldroubi A. (1996): "The Wavelet Transform: A Surfing Guide". pp.3-36, Wavelets in Medicine and Biology, Aldroubi, A., Unser, M., (ed.), CRC Press.
- Bovik A. C., Clark M., Geisler W. S. (1990): "Multichannel Texture Analysis using Localized Spatial Filters", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 12, s. 55-73.
- Chatterjee S. (1991): "Classification of Natural Textures using Gaussian Markov Random Field Models". pp.159-177, Markov Random Fields, Theory and Applications, Chellappa, R., Jain, A., (ed.), Academic Press.
- Du Buf J.M.H., Kardan M., Spann M. (1990): "Texture Feature Performance for Image Segmentation". Pattern Recognition, Vol. 23, No. 3/4, pp.291-309.
- Duda R.O., Hart P.E. (1973): "Pattern Classification and Scene Analysis". John Wiley & Sons.
- Haralick R. M. (1979): "Statistical and Structural Approaches to Texture". Proc. of the IEEE, Vol. 67, No. 5, pp.786-804.
- Jain A.K., Duin R.P.W., Mao J. (2000): "Statistical Pattern Recognition: A Review". IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, pp.4-37.
- Jernigan M.E., D'Astous F. (1984): "Entropy-based Texture Analysis in the Spatial Frequency Domain". IEEE Tran. Patt. Anal. Machine Intel., Vol. 6, pp. 237-243.
- Kil D.H., Shin F.B. (1996): "Pattern Recognition and Prediction with Applications to Signal Characterization". AIP Press.
- Li S.Z. (1995): "Markov Random Field Modeling in Computer Vision", Springer-Verlag.
- Manjunath B.S., Chellappa R. (1991): "Unsupervised Texture Segmentation using Markov Random Field Models". IEEE Tran. Patt. Anal. Machine Intel., Vol. 13, pp.478-482.
- Misiti M., Misiti Y., Oppenheim G., Poggi J.-M. (1996): "Wavelet Toolbox". The MathWorks Inc.
- Schürmann J. (1996): "Pattern Classification: A Unified View of Statistical and Neural Approaches". John Wiley & Sons.
- Van de Wouwer G., Scheunders P., Van Dyck D. (1999): "Statistical Texture Characterization from Discrete Wavelet Representation". IEEE Trans. On Image Processing, Vol. 8, pp. 592-598.
- Weszka J. S., Dyer C. R., Rozenfeld A. (1976): "A Comparative Study of Texture Measures for Terrain Classification". IEEE Trans. on Systems, Man, and Cybenetics, Vol. SMC-6, No. 6, pp. 269-285.