

**TIBBİ GÖRÜNTÜLEME TEKNİKLERİNİN ÜSTÜNLÜKLERİNİN  
KARŞILAŞTIRILMASINDA KULLANILAN BİR METODUN  
AKTUEL DURUMU :  
"Receiver Operating Characteristics (ROC)"\***

**Dr.Bülent Tekinsoy\*\*, Dr.Jörg Oestmann\*\*, Dr.Michael Galanski\*\*\***

Yeni tıbbi görüntüleme yöntemleri, bilinen eski görüntüleme teknikleri ile; radyolojik tanıda bir rol oynama açısından yarışma halindedir. Hem yeni hem eski yöntemlerin doğru yorumlanması gerek klinik gerekse ekonomik nedenler açısından büyük anlam taşımaktadır. Böyle bir yorumlamanın standartları bilimsel olarak değerlendirilebilmelidir. Birbirine benzer sistemlerin, örnek olarak çeşitli filmranforsatör sistemlerinin veya konvansiyonel ve digital radyografilerin verim farklılıkları araştırılmak istendiğinde bu standartların saptanması özellikle önem kazanmaktadır.

Görüntüleme sistemlerinin kıyaslanmasını içeren birçok araştırma, saptanan parametrelerin çok çeşitli olması güçlüğünden yakınmaktadır ve bu nedenle araştırılan yöntemin klinik teşhis konmasındaki anlamının açıklık kazanmadığı ve istatiksel olarak tekrar değerlendirilmesinin güç olduğu vurgulanmaktadır. Test edilen görüntüleme yönteminin değerlendirilmesinde kullanılacak "Goldstandard"ın saptanması için hiçbir uğraşmaya girilmemektedir (Burada sözü edilen Goldstandard, karşılaştırılan yöntemlerden üstünlüğü tartışmasız, referans bir yöntemi ifade etmektedir). Böyle incelemelerin sonuçlarının yorumlanması bu nedenlerden dolayı güçtür ve çok nadir olarak birbirleri ile kıyaslanabilirler. Buna karşılık bir "Receiver Operating Characteristics" (ROC)

\* Medizinische Hochschule Hannover (MHH, Hannover Tıp Fakültesi),  
1. Diagnostik Radyoloji Kliniği çalışmalarından.

\*\* MHH, 1. Diagnostik Radyoloji Kliniği Asistanı.

\*\*\* MHH, 1. Diagnostik Radyoloji Kliniği Profesörü.

araştırmasının yapısı basittir ve bu problemlerin çözümünü sağlamaktadır. Araştırmanın sonunda elde edilen değer, yani diagnostik verimlilik, test edilen görüntüleme yönteminin teşhise olan katkısını göstermektedir.

Bu yazıda ROC metodu bilimsel inceleme yapan klinisyenin bakış açısı ile açıklanmış ve ayrıca bizim karşılaştırmalı çalışmalarımızda gözlemlediğimiz bulgular örneklenmiştir.

#### SENSİTİVİTE; SPESİFİTE VE KİŞİSEL TAKDİRİN ROLÜ :

Her diagnostik yorum, gerçeğin (hastanın sağlık durumunun) şüphesiz, tartışmasız bilinmesi şartı ile şu dört kategoriden birine dahil edilebilir : Doğru-pozitif, doğru-negatif, yanlış-pozitif ve yanlış-negatif.

Diagnostik yorumların Doğru-pozitif kısmı tekniğin SENSİTİVİTESİ ve doğru-negatif kısmı ise SPESİFİTE'si olarak tanımlanır. Sensitivite ve spesifite çok yakın bir ilişki içerisindedirler ve sadece bir çift olarak değerlendirilmelidirler.

Eğer farklı iki görüntüleme sistemi, P ve Y, bu iki parametre dayanılarak karşılaştırılacak olursa, çizim 1'deki örnekte görüldüğü gibi her iki sistemin tanımlanması ortaya çıkar. Verilen parametrelerin uzun süre karşılaştırmalı incelenmelerine rağmen, hangi görüntüleme yönteminin üstün olduğu açıklık kazanmamaktadır. Bu açıklamayı yapamayışımız değişik iki etki ile karşı karşıya olduğumuz gerçeğine dayanmaktadır. Bunları;

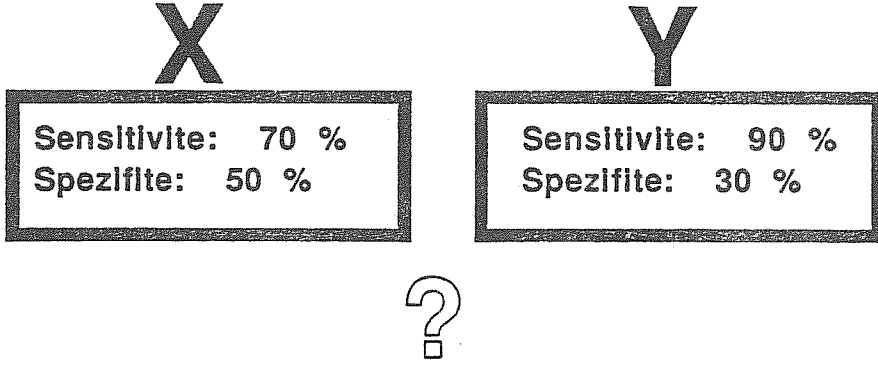
- A) Görüntüleme sisteminin vakaları "normal" ve "hasta" olarak gösterebilme yeteneği,
- B) Radyoloğun bireysel yorumu

şeklinde tanımlayabiliriz. Çizim 2 bu ilişkileri grafik olarak açıklamaktadır: Hem "normal" hem de "hastalık" kısımlarının dağılımları normaldir ve görüntüleme sisteminin verim kabiliyeti oranında birbirleriyle çıkmakta ve birbirlerinin belirli bir kısmını kapsamaktadırlar. Bu birbirlerini kapsayan ortak kısım ne kadar az bir alanı içerirse görüntüleme sisteminin diagnostik değeri o kadar fazladır.

Bu durumda bilimsel olarak aynı düzeydeki her radyolog, kişisel bir takdir eşğine ulaşmaktadır ve bu, çizimde dikey çizgilerle belirtilmiştir. Eğer bir tane "hasta"yı bile gözden kaçırmak istemeyen dikkatli, temkinli bir meslektaşımız (A) ile karşı karşıyaysak,

**Problem:**

Hangi görüntüleme yönteminin diagnostik değeri daha fazladır?



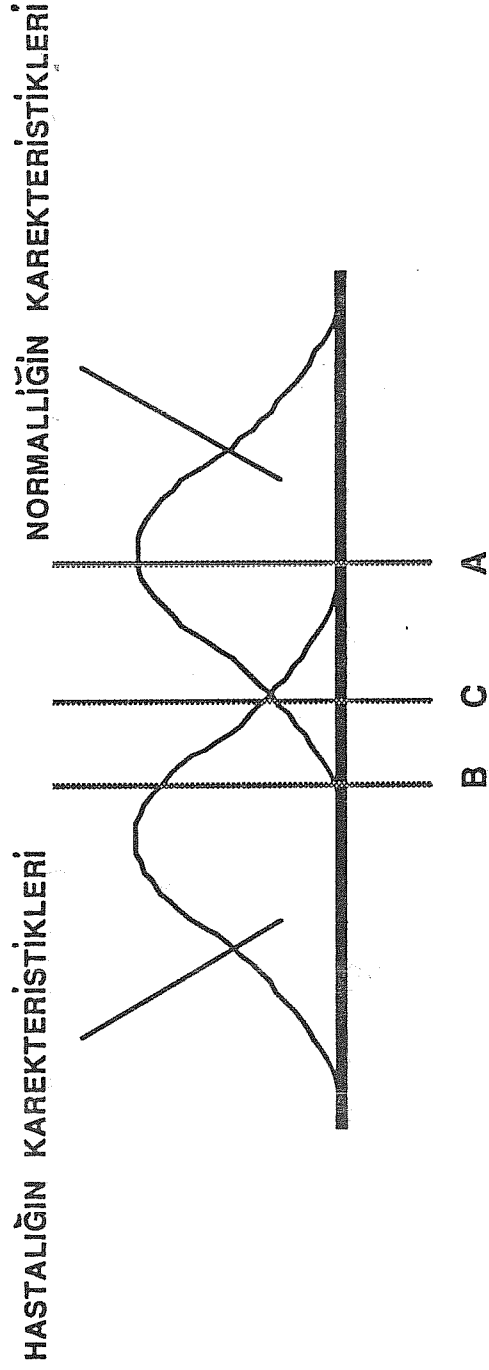
**Çizim 1.** Sensitivite ve Spesifitenin Saptanmasına Didaktik Bir Örnek.

bir çok "normal" durumu "hasta" olarak yorumladığını göreceğiz. Aksine eğer hiçbir "normal" durumu yanlışlıkla "hasta" olarak yorumlamak istemeyen bir meslektaşımız (B) ile muhattap isek onun bu görüşü nedeniyle büyük oranda gözden kaçmış "hasta" vakalarıyla karşı karşıya kalacağımızı bilmemiz gerekir. Normal yorumcu (C) ise tabii olarak bu iki ucun arasında herhangi bir noktada bulunacaktır. Hekimin bu farklı takdir eşiği, aynı görüntüleme sistemi içinde çeşitli sensitivite-spesifite çiftlerinin doğmasına neden olabilecektir.

İşte, ROC analizi bu zor durumun tam tersine bir başarı yaratmaktadır. Bunun içinde kişisel takdir eşiğini metodik olarak incelemektedir. Yorumcudan evet-hayır şeklinde cevaplar istenmeyip, teşhisinin kesinliğini kademeli bir tabelada kodlaması istenmektedir. Bu tabela en az 3 kademeli olmalıdır. ROC analiz yöntemini kullanarak kendi kliniğimizde yaptığımız bir çalışmayı burada örneklemek, tekniği anlama ve yorumlama bakımından daha yararlı olacaktır. Bu amaçla biz pulmoner nodülleri içeren bir araştırmada 5 kademeli bir skala kullandık. Bu skalada yer alan şıklar şunlardı:

- 1) Kesinlikle pulmoner nodül yok
- 2) Muhtemelen pulmoner nodül yok
- 3) Yorum mümkün değil
- 4) Muhtemel pulmoner nodül
- 5) Kesin pulmoner nodül

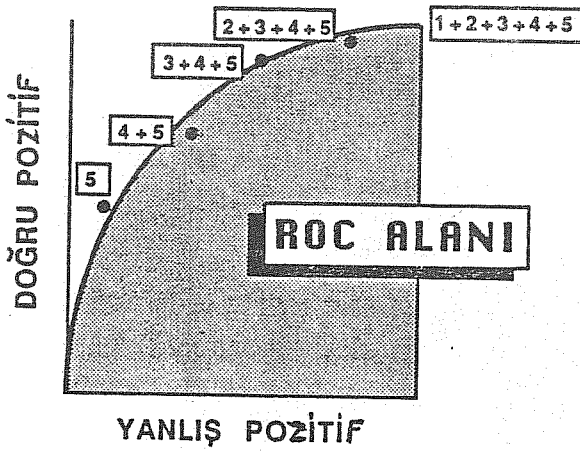
## KARAR EŐİKLERİ - CONFİDENCE -



Çizim 2. Verimin Yorumlanmasına Etkisi Olan Faktörler (Açıklama Yazının İçerisinde).

Saptanan teşhisin böyle kademelendirilmesi spesifite-sensitivite çiftlerinin oluşturulmasını (yukarıdaki durumda 5 tane) sağlar. 1'den 5'e kadar sıralanmış bu farklı basamakların herbiri için yapılan doğru pozitif ve yanlış pozitif yorumların doğrudan toplanması sonuçta "sıfır noktası" da dahil olmak üzere ROC eğrisini ve bu eğrinin altında kalan, diagnostik verimi ifade eden alanı verir (Çizim 3)<sup>3</sup>. Çizim 1'de verilen örneğin aksine test edilen her görüntüleme sistemi için sonuç olarak tek bir sayısal değer elde edilecektir ve bu sayede sistemlerin kalitatif sınıflanması mümkün olacaktır: X yöntemi diagnostik verim olarak 0.95 değerine sahiptir (ROC alanı) ve diagnostik değeri 0.75 olan Y sistemine üstünlük sağlamaktadır.

## "DIAGNOSTİK VERİM"



Çizim 3.  
ROC Eğrisinin ve Alanının Saptanmasında Kullanılan Koordinatlar (Açıklama yazıda).

## ARAŞTIRMANIN ANA YAPISI

İki görüntüleme yönteminin karşılaştırılabilmesi, kıyaslanabilmesi için ya aynı vaka topluluğu her iki sistem ile incelenmelidir veya birbirine benzeyen iki ayrı vaka topluluğu sistemlerin sadece biri ile araştırılmalıdır. İlk metod elde edilen verilerin karşılaştırılabilirliğini garanti etmektedir. Ancak çeşitli incelemeler arasında büyük zaman farkı olmaması gerekir. Seçilen vakaların % 50'si "hasta", % 50'si "normal" olmalıdır. Bu sayede doğru pozitif ile yanlış pozitif teşhislere aynı oranda fırsat tanınmış olur. Değerlendirmeyi yapacak olanlara, vakalar rastgele seçilerek verilmeli-

dir. Vaka sayısının düşük olduğu durumlarda, her yeni araştırmanın başında bu sıraların değiştirilmesi vakaların hatırlanabilmesini güçleştirecektir. Yine aynı nedenden dolayı araştırmalar arasında yeterli zaman aralığının bırakılması gerekir. Sistematik öğrenme etkisini kompanse etmek için değerlendirmeyi yapanlara karşılaştırılan görüntüleme yöntemlerinin görüntü birimleri (film grupları) değişik şekillerde sıralandırılarak sunulmalıdır. ROC araştırmasında sık karşılaşılan sorunların üzerinde durmak gerekir.

### GERÇEĞİN SAPTANMASI

Görüntüleme yöntemlerinin sistematik olarak kıyaslanabilmesi için "hastalık" ve "normallik" (bulgunun olması veya olmaması) açık olarak araştırmanın başında saptanmış olmalıdır.

Fantom araştırmalarda tabii olarak bu şartı yerine getirmek çok kolaydır. Tamamiyle bir fantom araştırmayı içeren incelemede dahi klinik uyum kesinlik kazanmalıdır: Araştırmayla ilgili lezyonlar ile normal anatomik yapıların yeterli düzeyde gösterilmiş olması gereklidir<sup>2,3</sup>. Gerçeğe yakın fantomların oluşturulmasına çalışılmalıdır. Bu güçlük, gönüllü adayların kullanılması ve istenilen lezyonların bu görüntülere oturtulmasıyla kısmen azaltılmış olur. Bu metotta da oluşturulan yapay lezyonlar ile gönüllülerin "normalliği" yeterli düzeyde birbiriyle uygun olmalıdır.

Klinik hasta vakalar grubu incelemeye dahil olduğunda ve gönüllüler normal grup olarak kabul edildiğinde, hem patolojinin hem de normalliğin ispatı gerekir. Bu durumda bir metodun "Goldstandard" olduğu görüşünde birleşilmelidir. Bu metodun diagnostik güvenliğinin tartışmasız veya en azından karşılaştırılan yöntemlerden çok üstün olması gereklidir.

Bir hastalığın (yer kaplayan lezyon vs.) tartışma götürmeyecek şekilde kesin ispatında kural olarak histopatolojik ve cerrahi bulgular gereklidir. Bazı hallerde ise belirgin klinik bulgular veya özgün laboratuvar değerleri bir hastalığın kesin tanımı için yeterli olabilirler.

Normalliğin ispatı ise daha güçtür. Bir akciğer filminde neoplastik bir gelişimin olmadığı ancak klinik olarak yeterli sayılacak belirli bir zaman aralığından sonra tekrarlanan kontrol filminin bir özellik göstermemesi ile kabul edilebilir. Belirgin olarak üstün olan bir görüntüleme sistemi (mesela bilgisayarlı tomografi) teşhis için yeterli bilgi vermekten uzak olan görüntüleme yöntemlerinin

karşılaştırılmasında goldstandard olarak kabul edilebilir. Normal ve patolojik tanının bir uzmanlar kurulunca (experten panel) incelenen görüntüleme yöntemleri aracılığı ile araştırılması gerçeğin saptanmasında kullanılacak zayıf metodlardan biridir. Hangi goldstandard seçilirse seçilsin; o metoddan daha üstün bir metod olmadığı kabul edilmelidir.

## SAPTANAN PATOLOJİNİN BELİRGİNLİĞİ (GÖZE ÇARPICILIĞI)

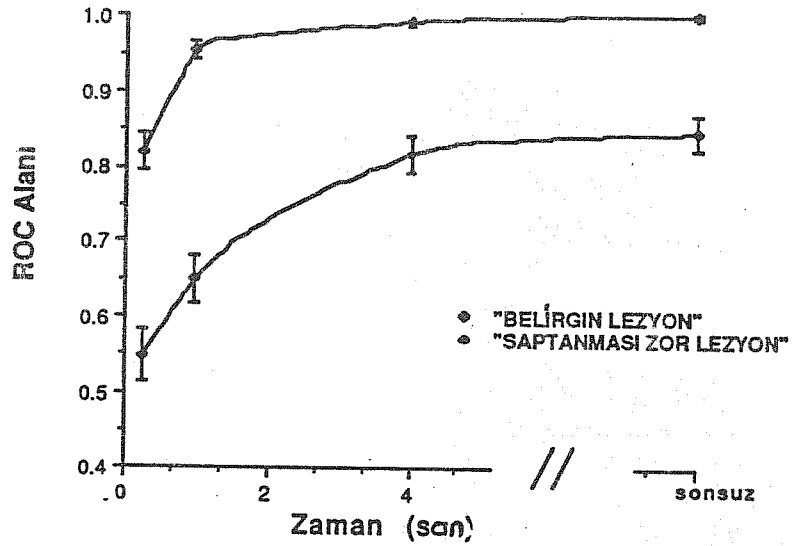
Bir lezyonun göze çarpıcılığı; büyüklüğüne, çevresine oranla sahip olduğu kontrasta ve dış sınırlarının özelliğine bağlıdır. Seçilen lezyonların belirginliği farklı olmalıdır. Bunun nedeni öncelikle gerçek klinik patolojinin çok yönlü olduğunun gösterilmek istenmesidir ve yorumcunun ROC'nin kademeli skalasının farklı eşiklerini kullanabilmesi sağlamaya çalışılmaktadır.

Belirginliğin oranı, incelenilen yöntemlere göre saptanmalıdır. Yöntemlerin boyutlarının saptanmasında seçilecek lezyonların büyük çoğunluğu incelenilen yöntemler ile farkedilebilirliğin sınırında olmalıdır. Lezyonların tümü, ne karşılaştırılan yöntemlerin hiçbiri ile saptanamayacak kadar belirsiz, ne de tüm yöntemler ile gözden kaçırılmayacak kadar belirgin olmalıdır. Aşırı belirgin bulgularla dolu ROC analizi, mükemmel diagnostik verim sonucunu uyandıran yanlış eğrilerin oluşmasına ve bu nedenle yöntemler arasındaki (anlamlı) belirgin farkların saptanmasının engellenmesine neden olur. Gerekli belirginlik boyutu tahmin edilemiyorsa, küçük bir pilot araştırmanın yapılması tavsiye edilmektedir. Eğer mümkünse, belirginlik boyutunu sonuçta 0.65 ile 0.95 arasında yer alan bir ROC alanı elde edilecek şekilde kalibre etmeye çalışılmalıdır.

Klinik ROC incelemelerinde belirginliği az olan lezyonların bulunduğu vakaları yakalamak genellikle zordur. Bu güçlüğü'n nedeni, ideal bir ROC araştırmasında normal ve patolojik vakaların % 50 oranında temsil edilmeleri gerekliliğine dayanmaktadır. Bu durumda pozitif bir bulgu ihtimali, kural olarak, klinik rutine oranla çok fazladır. Böylece klinik rutinde sıklıkla gözden kaçan bir lezyon, ROC araştırması içinde yapay olarak oluşturulan yüksek hastalık ihtimali ve yönlendirilmiş deteksiyon (saptama) uyarısı ile, belirgin bir sinyal haline gelebilir. Bunun dışında az göze çarpan lezyonların teşhis edilmeleri tabii olarak daha nadirdir ve sıklıkla daha güç olarak tanımlanırlar. Eğer klinik bir pilot incelemede elde edilen lezyonlar çok büyük ROC alanlarının ( $> 0.95$ ) oluşmasına neden olurlarsa, inceleme süresinin kısaltılması (mesela

bir PA toraks filmi için birkaç saniye verilmesi) alanın küçültülmesini sağlar<sup>4</sup>. Bu ayrıca yöntemin hassasiyetini arttırır (Çizim 4).

## FARKEDİLEBİLİRLİK VE İNCELEYİCİYE TANINAN SÜRE



Çizim 4. Yorumcuya Tanınan Sürenin Kısaltılmasıyla Diagnostik Verimde Düşmektedir. Aynı Nedenle Lezyonların Belirginliği ile Diagnostik Verim birbirleriyle bağlantılıdır.

## LEZYONLARIN LOKALİZASYONU

Tipik bir ROC incelemesinin ana yapısı dahilinde, sadece lezyonların olup olmadığı sorulur, lokalizasyonları ise sorulmaz. Bunun sonucu olarak inceleyen, gerçek lezyon yerine başka bir yapıyı ifade edebilir ve buna rağmen sonuçta doğru pozitif sonuca ulaşabilir. Bu nedenden dolayı Localisation Receiver Operating Characteristics (LROC) analizi geliştirilmiş, ancak bu metod için bu güne kadar güvenilir bir anlamlılık hesaplama metodu tanımlanmadığından bu metoddan da vazgeçilmiştir<sup>5</sup>. Saptanan lezyonun lokalizasyonu, ROC kademe eşliğinden bağımsız olarak sorulabilir ve sonuç olarak "accuracy" veya teşhisin doğruluğu şeklinde ayrı



olarak açıklanabilir (Çizim 5) <sup>6-8</sup>.

## KONVENSİYONEL VE DİĞİTAL MAMMOGRAFINİN DİAGNOSTİK VERİMİ VE TEŞHİS DOĞRULUĞU

| Görüntü Sistemi  | ROC Alanı   | Accuracy % |
|------------------|-------------|------------|
| Film-Ranforsitör | 0.87 ± 0.06 | 61         |
| Depo-Fosfor      | 0.87 ± 0.04 | 62         |

Çizim 5. Vaka örnekleri: ROC alanı ve Accuracy (teşhis doğruluğu), test edilen sistemler için belirtilmiş ve bir kıyaslama yapılmaya çalışılmıştır.<sup>6</sup>

### İNCELENECEK VAKA SAYISI (KOLLEKTİF BÜYÜKLÜĞÜ)

Esas olarak kollektifin büyüklüğü, kıyaslanılan yöntemlerin diagnostik verimlerinin tahmin edilen farklılığına bağlıdır. Tahmin edilen verim farkı ne kadar az olursa, o oranda daha fazla vaka incelenmelidir. Doğal olarak, oluşan farklılıkları en belirgin şekilde ortaya koyacak bir araştırma grubu oluşturmaya çalışılmaktadır. Küçük bir pilot inceleme, araştırılan yöntemlerin geçici ROC alanlarının gösterilmesini mümkün kılar. Özel tabelalar yardımı ile daha sonra gerekli vaka sayısı saptanabilir. Tereddütlü olduğu tartışmalı hallerde daima yüksek rakamlar kabul edilmelidir. Vaka sayısının 100'ün (50 normal, 50 hasta) altında olduğu durumlarda dikkatli olunmalıdır<sup>6</sup>. Vaka sayısı, bir görüntüleme sisteminin tek tek her filminin bir vaka gibi kabul edilmesiyle kendi sınırları içerisinde genişletilmiş olur. Böylece CT ve NMR incelemelerinin her kesitleri için tek tek gerçek saptaması mümkünse, her biri bir vaka gibi incelemeye dahil edilebilirler.

Projeksiyon radyografilerinin yorumlanmasını içeren araştırmalarda, gözlem sayısının arttırılması, görüntülenen bölümün ayrı ayrı kısımlara ayrılması ile mümkün olabilir.

Gözlemci sayısının yüksek tutulması ile düşük vaka sayısı dengelenemez.

### GÖZLEMCİNİN BİLGİ DÜZEYİ

Gözlemcinin (bilgi) seviyesi, araştırmanın cevaplamaya çalıştığı soruların düzeyine uygun olmalıdır. Görüntüleme yöntemleri ile elde edilen diagnostik verimlerin incelenmesinde kural olarak gözlemciler, radyologlar olmalıdır. Gerçeği uygun olmayan fantomlarla yürütülen araştırmalarda gözlemcinin kabiliyeti o kadar önemli değildir. Yeni ve az alışılmış görüntüleme yöntemlerini içeren araştırmalarda, gözlemcilere önceden çalışma imkanı verilmelidir.

### ANLAMLILIK (SİGNİFİKANZ) TESPİTİ VE GEREKLİ GÖZLEMCİ SAYISI

Görüntüleme yöntemlerinin diagnostik verimlerinin farklılıkları, bazı nadir hallerde, sadece ROC eğrilerinin gözlenmesiyle ortaya konabilir. ROC analizleri için kullanılan bilgisayar programlarında kural olarak ROC alanlarının standart sapmaları belirtilmiştir. Bireysel olarak katılan gözlemcilerde daha doğru bir anlamlılık analizi için bir bilgisayar programı hazır bulunmaktadır.

Pratikte araştırmacıya bağımlı olarak T-test çifti sabit olarak korunmuştur. Bu test çifti ile her gözlemcinin araştırılan her görüntüleme yöntemi için elde ettiği ROC alanı kıyaslanabilmektedir. Eğer anlamlılık hesaplanmasında bu şekilde bir inceleme yürütülmesi için çaba sarfedilirse, gözlemci sayısının (6 veya daha fazla) yeterli düzeyde yüksek olması gereklidir ve ROC alanı değerlerinin normal dağılımı elde edilmelidir. Aksi halde parametrik olmayan bir yöntem, mesela Wilcoxon testinin uygulanması gereklidir.

Eğer ikiden fazla görüntüleme yöntemi birbiri ile karşılaştırılıyorsa arzu edilen T-test anlamlılık düzeyi Bonferroni yöntemine göre düzeltilmelidir. Bu yöntemde elde edilen değerler karşılaştırılan yöntem sayısına bölünmektedir.

### SONUÇ

ROC analizi, bilimsel olarak açıklanmış bir yöntem olup, diagnostik sistemlerin kıyaslanmasında kullanılan pratik ve duyarlı

bir metoddur. Bazı ana kurallara sadık kalındığında kullanımı, uygulanması basittir. Modern teknik ve büyük masraf gerektiren yeni görüntüleme yöntemlerinin doğru yorumlanmasında, haklarında doğru bir kanıya varılmasında değerli bir araçtır.

## KAYNAKLAR

1. Godfrey K. Comparing the Means of Several Groups. *N Eng J Med* 313: 1450-1456, 1985.
2. Hanley JA, McNeil BJ. The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve. *Radiology* 143: 29-36, 1982.
3. Kundel HL, Nodine CF, Thickman D, Carmody D, Toto L. Nodule Detection With and Without a Chest Image. *Invest Radiol* 20: 94-99, 1985.
4. Metz CE. Basic Principles of ROC Analysis. *Seminars in Nuclear Medicine* 8(4): 283-298, 1978.
5. Oestmann JW, Greene R, Kushner D, Bourgouin PM, Linetsky L, Llevellyn HJ. Viewing Time and Detectability of Lung Lesions. *Radiology* 166: 451-453, 1988.
6. Oestmann JW, Kopans D, Hall DA, McCarthy KA, Rubens JR, Greene R. A Comparison of Digitized Storage Phosphors and Conventional Mammography in the Detection of Malignant Microcalcifications. *Investigations Radiology* 23(10): 725-758, 1988.
7. Oestmann JW, Kopans D, Linetsky L, Hall DA, McCarthy KA, Swann C, Kelley JE, Johnson L. Comparison of Two Screen Film Combinations for Mammographic Studies. *Radiology* 168: 657-659, 1988.
8. Revesz G, Kundel HL, Graber MA. The Influence of Structured Noise on the Detection of Radiologic Abnormalities. *Invest Radiol* 9: 479-486, 1974.
9. Starr SJ, Metz CE, Lusted LB, et al. Visual detection and localisation of Radiographic Images. *Radiology* 116: 533, 1975.

