

ROC (receiver operating characteristic) 해석

조선대학교 치과대학 구강악안면방사선학교실

김재덕

Interpretation of Receiver Operating Characteristics (ROC)

Jae-Duk Kim

Dept. of Oral and Maxillofacial Radiology, School of Dentistry, Chosun University

ABSTRACT

The purpose of this paper is to explain the making procedure and the usage of receiver operating characteristic (ROC) curve for interpretation of radiographic images. The conventional radiograms obtained after the creation of the lesions in the acrylic plates and were enhanced in color. The observer were informed of which tooth to examine, the 'a priori' probability of a lesion present and the approximate diameter of the lesions. The two groups of films were interpreted separately by the same observer using the same rating scale. The following rating scale was used: A; definitely no lesion, B; probably no lesion, C; not sure, D; probably a lesion, and E; definitely a lesion. In analysis, for each observer the diagnostic results in terms of true positive (TP) and false positive (FP) decisions were plotted on a graph. The lowest point on the graph represents the TP and FP when only decisions designated as E according to the rating scale are included. The next point shows the TP and FP values when diagnoses designated as D are added and so forth. By connecting such plot points, a receiver operating characteristic (ROC) curves is obtained. The area under the curve represents the diagnostic accuracy resulting from a diagnostic performance at pure chance level and a value of 1.0 at perfect performance. This method has been known as an useful method to detect the minute difference for each radiographic technic, each observer and for the different lesion depths. (*Korean J Oral Maxillofac Radiol 2000 ; 30 : 155-158*)

KEY WORDS : interpretation of film images, ROC curve, diagnostic accuracy

서론

ROC해석법 (Mets, C.E. 1979)¹은 accuracy, sensitivity, specificity 등의 진단에 있어 질환의 판정에 대한 평가방법의 결점을 보완하기 위해 고안된 평가방법이다. 많은 화상 시스템의 물리적특성은 공간해상력, noise의 크기와 구조 (Wiener spectrum), 대조도 (특성곡선)에 의해 나타내어진다. 그러나 화상으로부터 얻어지는 일반 화상의 물리적 특성을 산출해 내었다 하여도 질병에 대한 판단은 최종적으로 결정될 수 없다. 현재의 진단행위는 관찰자가 각각 화상이 얻어졌을 때 피사체의 상태에 관해 각자의 판단을 기록함에 의해 경험적으로 결정된다. 인간 시각의 복잡한 체계는

잘 알려지지 않기 때문에 관찰자의 능력을 인정할 수 있도록 시각에 있어서 검출과 인지에 대한 신뢰도가 높은 모델의 개발이 필요하다. 그렇다면 어떻게 하여 관찰자능력이 측정될 수 있을까?

ROC해석법²은 두 가지 화상의 미세한 차이가 관찰자의 관점에서 평가될 수 있으므로 연구평가에 이용될 수 있다. 여러가지 목적으로 임상에서 영상이 만들어지고 있는바, 정상조직 중에서 이상이 있는 음영을 검출하는 작업이 그 중 하나라고 말할 수 있다. 병변을 신호라고 생각할 때, 이 신호를 효율적으로 검출할 수 있는 촬영시스템을 구하려는 연구들은 상당히 의미있는 일로서, 특히 시각에 의존하여 평가하는 방법에 대한 실험이 보다 더 진척되면 실제 임상에 보다 가까운 결과를 기대할 수 있다고 생각된다.

ROC 곡선은 상당히 미미한 신호를 사람의 시각에 의해 검출되는 정도를 계산하는 방법으로서는 우수한 편이어서 현재까지 많은 성과를 올리고 있다. 그러나 ROC 곡선은 약신호에만 적용하는 것이 가능한 방법으로 그 검출이 가능

접수일: 2000년 5월 6일 채택일: 2000년 6월 2일
Correspondence to : Prof. Jae-Duk Kim
Department of Oral and Maxillofacial Radiology, School of Dentistry, Chosun University, 588 Susuk-dong, Dong-gu, Kwangju, KOREA
Tel) 062-220-3633, Fax) 062-227-0270
E-mail) jdkim@mail.chosun.ac.kr

할까, 가능치 못할까를 예측의 대상으로 하고 있는 점에 주의해야 할 필요가 있다. 즉 신호의 형태, 질, 크기와 성상을 평가하는 것은 가능하지 않음을 알아두는 것이 필요하다.

시료의 준비

무엇을 평가하는것일까, ROC 곡선은 적응가능할까

이상과 같은 주의사항을 염두에 두고 이제부터 시작하려는 연구에 적당한 평가방법인지 아닌지를 잘 살필 필요가 있다. 연구방향을 결정한 다음 그 문제점에 대한 분석은 상당히 중요하다. 먼저 관찰할 시료를 준비한다.

무엇을 평가하는 것인가: 어떤 형태의 결과를 얻고 싶은 것인지를 충분히 검토한 후 평가에 적용되는 시료를 준비한다. 여기서 예로 거론될 실험에서는 신호의 검출에 신호의 크기가 어떠한 형(어떤 형태의 사진)에서 영향을 받고 있는 것인지를 평가하기 위한 기초적인 검토를 시행하는 것을 목적으로 하고 있으므로,

직경 2mm의 아크릭비즈를 촬영한 일반방사선사진 시료와, 또하나 칼라화한 방사선사진 시료를 만들었다.

여기서 주의하지 않으면 안될 것은 신호가 확실히 보일 것 같은 시료라면 ROC 곡선에 의한 평가는 할 수 없다고 말할 수 있다. 적절한 시료를 작성하기 위해서는 여러 번의 시도와 관찰의 조절을 필요로 한다. 혹시 신호가 강하기만하든지, 적당히 약한 신호가 얻어지기 어려운 때는 신호의 감약(예를 들어 산란체를 첨가)도 생각하지 않으면 안된다.

최적의 관찰시료를 만드는 것이 가장 중요하다. 시료의 수는 통계적으로 모순이 없는 수가 필요하다. 구체적으로는 하나의 계에 50-100군의 신호가 있는 시료와, 없는 시료를 준비하면 좋다고 말할 수 있다. 다만 임상사진을 이용하는 경우는 이러한 수의 시료를 준비하는 것이 상당히 곤란하므로 꽤 적은 시료에서 평가하지 않으면 안되는 경우도 있다.

관찰 및 판정²

1. 관찰

관찰에 임하여는 다음과 같은 준비가 일반적으로 이루어져야 한다.

1) 관찰자의 선정

일반적으로 방사선기사와 의사가 관찰자가 되는 경우가 많은 것 같다. 목적에 부합되는 인선을 해야 하며 실험의 목적 및 의의에 대해 이해와 협력이 절대 조건으로 필요하다. 그를 위해서 충분한 토론을 해야 한다.

2) 관찰환경의 설정

X선 필름을 이용한 시료를 관찰하는 경우는 관찰자의 피로를 경감하기 위해 시료의 크기에 맞는 창 이외의 부분은 모두 차광하는 등의 배려를 한다. X선 필름이외의 시료, 예를 들어 CRT 화면의 경우에도 각각 보기에 좋게 배려를 한다. 관찰거리, 시간, 주위의 밝기, 소음 등 관찰자의 심리에 영향을 미칠 수 있는 외적인자는 실험의 기간 중 가능한 일정하게 할 필요가 있다. 예비실험에서 어느 정도가 좋은지를 확인할 필요가 있다.

3) 관찰방법의 결정

관찰방법에는 여러 종류가 보고되어있는바 실험에 맞는 방법을 선택하면 좋겠다. 평정법에서는 5단계 평가가 일반적이나 최근에는 무한의 단계를 기본으로 한 연속확신도법이라 하는 것도 개발되어있다. 여기서는 5단계 평점을 예로 하겠다. 읽어내는 순서효과는 동일한 관찰자에서 2개 이상의 계를 비교하려는 경우에 먼저 시행한 관찰의 정보를 그 다음의 관찰 시에 갖고 넘어가는 일이 발생한다. 이를 방지하기 위해 관찰시간을 충분히(1-2주간)했다가 관찰순서를 바꾸었다하는 방법을 취한다. 관찰방법은 5단계 평정을 이용하였으나 관찰방법으로서는 참고로 아래와 같은 방법이 있다.

Yes-No : 이것은 관찰자 자신이 판정기준을 변화시켜가며 검출을 행한다(엄격히 보면 신호는 있으나 약간 기준을 채우지 못하면 신호는 없다는 식으로). 실제 응용 곤란한 방법이지만 신호검출이론의 기초적 방법이다.

판정 수정기 : 5단계 정도의 판정기준을 확실히 설정해 두고, 관찰자는 그 판정기준에 따라 판단한다. ROC 곡선을 작성하는 일반적 방법이다.

강제선택법 : 관찰자는 신호가 있는 화상과 얼마간의 신호가 없는 화상을 동시에 관찰하여 그 속에서 어떤 화상에 신호가 있는가를 답한다. 2지 강제선택법(2AFC), 18지 강제선택법(18AFC)이 있다. 여기서는 5단계의 판정단계를 이용한 평정방법으로 실험을 시행하였으나 문제는 없다고 말할 수 있다.

4) 이론적 배경의 검증

실험을 진행하는 때는 무엇을 평가하려고 있는가, 평가방법은 아주 틀린 것이 아닐까, 데이터의 기초가 되는 시료는 내용이나 수가 목적에 맞는가, 수순은 올바른가, 관찰자의 선택에 문제는 없는가, 결과의 정리나 처리방법은 좋은가, 등등 실험의 그때그때 체크해나가는 작업이 필요하다. 이 같은 이론적 배경은 실험 시작 전에 충분히 검토하지 않으면 안되나 작업의 도중에도 체크하는 것은 되돌아가는 길을 피하기 위해서도 필요한 일이다.

예로 하고 있는 실험은 신호의 크기가 관찰자의 신호 검출능에 미치는 영향을 조사하는 것이 목적이다. 이 실험에 ROC 곡선을 이용하고자 했다. 여기서 약간 이론적인 점을 서술하도록 해 실험이론의 검증을 하고자 한다. ROC

곡선은 “수신자 동작특성곡선”이라고도 말해져, 관찰자의 ‘자극에 대한 반응’을 나타내는 것을 알 수 있다. 횡축에는 신호가 들어있지 않은 시료를 본 경우 신호가 있다고 틀리게 답한 조건확률(오버확률): $P(XI)$ 을 취하고, 종축은 신호가 들어있는 시료를 보고 신호가 있다고 답한 조건 확률(적중확률): $P(XIs)$ 를 취한다. 여기서 ROC곡선이 생긴다. 이 방법은 신호검출이론을 기초로 한 것으로서 잡음의 속에서 약한 신호를 검출하려 할 때 이용되고, 제시된 예와 같은 실험에 이용하는 방법으로서 타당하다고 말할 수 있다.

2. 판정

시료가 언어적 실험의 준비가 정리되어 실제의 평가작업에 들어간다. 관찰자는 여기서 처음으로 시료를 보게되지만 판정하는 목표를 알지 못하면 바른 결과가 얻어지지 않는다. 그를 위해 사진에 신호의 크기, 어느 정도의 형, 화상의 속에 존재하는 양태, 수, 관찰방법(판정단계속지), 결과의 기록 등 판정에 필요한 내용을 철저히 훈련시킨 후 시행한다. 관찰방법을 충분히 이해할 수 있게 되었을 때 관찰에 들어간다. 제시된 시료에 대하여 관찰자는 주관적 판정을 5단계의 판정기준 하에서 행한다.

판정기준

- A. 신호가 없다. B. 신호가 없는 것 같다. C. 모르겠다.
- D. 신호가 있는 것 같다. E. 신호가 있다.

제시한 시료에 신호가 있다, 없나는 알고있으므로 신호가 있는 시료 50개, 신호가 없는 시료 50개 각각에 대하여 관찰자가 대답한 판정을 판정기준 A-E에 카운트한다. 실제에는 각 판정기준에 대답을 분류하는데 확실히 알고있는 시료에 나란히 대비하여 정확히 카운트하도록 주의한다.

여기서 $T(XIs)$ 는 신호가 들어있는 시료(noise+신호)의 판정기준 쪽의 카운트 수.

$T(XI)$ 은 신호가 없는 시료의 판정기준 쪽의 카운트 수이다.

여기서 관찰자의 신호에 대한 응답(ROC곡선)을 구하는 데에는 최종적으로 신호가 들어있는 시료를 본 경우 신호가 있다고 답한 조건에 대한 확률: $P(XIs)$ 와, 신호가 들어있지 않은 시료를 본 경우 신호가 있다고 잘못 답한 조건에 대한 확률: $P(XI)$ 을 얻지 않으면 안된다.

이를 위해 시료에 대한 답을 할 때 판정기준을 상기의 5단계로 설계하여 판단에 엄격한 단계를 붙여 분류했다.

판정결과처리

직경 2mm의 아크리비즈를 촬영한 일반방사선사진 시료와 또 하나 칼라화한 방사선사진 시료중 일반방사선사진 시료의 경우 각 단계에서 그 결과가 이하에 표시된 $T(XIs)$ 와 $T(XI)$ 이다.

이어 이번에는 신호가 없다고 대답한 (A)를 1로하여 각 판정기준에 대한 확률을 구하기 위해 단계 E부터 A의 방

시료 (일반방사선사진)	A	B	C	D	E
$T(XIs)$; noise+신호	9	11	13	9	8
$T(XI)$; noise	11	15	10	12	2

향으로 누적한 수를 각각의 판정기준에 기록한다.

이것을 $T(Xs)$: noise+신호, $T(X)$: noise로 하여 아래와 같이 나타내었다.

시료 (일반방사선사진)	A	B	C	D	E
$T(Xs)$; noise+신호	50	41	30	17	8
$T(X)$; noise	50	39	24	14	2

여기서 부여된 시료(임상사진의 경우도 있다)로부터 어느 정도의 확률(P)에서 목적으로 한 신호(혹은 병변, 변화 등)가 검출되었다는 것이 수치로서 표현 가능하게 된다. 아래에 그 결과가 있다.

신호가 있는 시료 $P(XIs)$	1.00	0.82	0.60	0.34	0.16	종축 TPF
신호가 없는 시료 $P(XI)$	1.00	0.78	0.48	0.28	0.04	횡축 FPF
	A	B	C	D	E	

확률 $P(XIs)$ 를 종축에 $P(XI)$ 을 횡축으로 하여 각 점을 plot하여 그래프를 만들면 일반방사선사진 시료에 대한 ROC곡선이 생긴다.³

이어 칼라화한 방사선사진시료에 대해서도 같은 식으로 실험함에 의해 두 개의 ROC곡선이 완성된다.

평 가

언어진 ROC곡선은 횡축에 신호가 들어있지 않은 시료를 보고 신호가 있다고 틀리게 답한 조건의 확률: $P(XI) = FPF$ 를, 종축에 신호가 들어있는 시료를 보고 시료가 있다고 답한 조건쪽 확률: $P(XIs) = TPF$ 를 입력하여 플롯하였기 때문에 그래프의 좌상방, 즉 각 판정기준에 있어서 $P(XI)$ 이 보다 0에, $P(XIs)$ 가 보다 1에 가까운 결과치가 검출이 좋은 것 즉 곡선하면적이 큰 것이 검출이 좋은 것으로 평가 된다.¹

이와같이 ROC 곡선을 그려보면 설득력이 있는 형태로 화상의 평가가 가능하고 그 표현도 상당히 간결하여 이해하기 쉬움을 알 수 있다.

한편, ROC곡선평가법은 확률적인 배경을 가진 평가법이므로 결과에 대해서 수치를 갖고 구체적으로 표현하는 방법이 몇 가지 이용되고 있다. 특히 복수의 ROC곡선이 유의한 차이가 있는건가, 어떤가를 판단하기 어려운 경우에

ROC 해석

참고할 수 있는 데이터를 제공하기 위한 방법으로 곡선하면적 (Accuracy): 통계적 유의차 검정, 감도 (Sensitivity), 특이도 (Specificity) 등 실험에 맞는 평가방법을 선택할 필요가 있다.

여기서

$$\text{sensitivity} = \frac{\text{No. of true positive decisions}}{\text{No. of actually positive cases}}$$

$$\text{specificity} = \frac{\text{No. of true negative decisions}}{\text{No. of actually negative cases}}$$

$$\text{accuracy} = \frac{\text{No. of true posi deci} + \text{No. of true nega deci}}{\text{No. of all cases}}$$

로 정의되며⁴ ROC곡선에서 커브아래 총면적이 diagnostic accuracy를 의미한다.^{5,6}

요 약

1. 일반방사선사진과 칼라화한 방사선사진의 비교에서

각각 필름에서 진단을 시행할 때 ROC해석법에서는 true positive fraction (TPF), false positive fraction (FPF)를 매개변수로 하고 있으므로 우선 두가지 필름형태에 대해 각각 따로 다음과 같이 평가한다.

2. 판정기준

병변없다 A, 거의없다 B, 모르겠다 C, 거의 있다 D, 있다 E

먼저 일반방사선사진에서 실제로 병소가 총있는 것이 50, 총없는 것이 50 인데 위 판정기준 각각에 대해

	A	B	C	D	E
실제있는 것	9	11	13	9	8 ; 총 50cases
실제없는 것	11	15	10	12	2 ; 총 50cases

TPF	50/50	41/50	30/50	17/50	8/50
	1	0.82	0.60	0.34	0.16 →1
FPF	50/50	39/50	24/50	14/50	2/50
1-p	1	0.78	0.48	0.28	0.04 →0

3. 곡선만들기

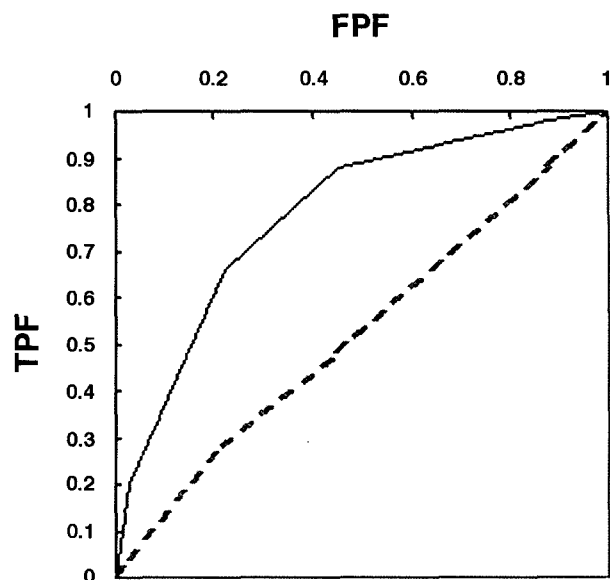
a. 횡축은 FPF 종축은 TPF로 한 그래프를 plot를 한다.

sensitivity 17/50 specificity 26/50 accuracy 43/100

b. 곡선만들기 프로그램을 이용하여 곡선을 만들시에는 TPF를 a에 입력하고 FPF를 b에 입력한다.

이 plot을 그릴 수 있는 프로그램은 <http://www.members.tripod.co.kr/jdakim> 또는 <http://www.chosun.ac.kr/~jdkim>의 홈페이지내 공개자료실에서 다운 받으실 수 있습니다.

Conventional Binormal ROC Curves



이프로그램에서 입력할 a, b의 값은

$$\text{TPF}^5 ; a = \frac{\text{No. of true positive decisions}}{\text{No. of actually positive cases}}$$

$$\text{FPF}^5 ; b = \frac{\text{No. of false positive decisions}}{\text{No. of actually negative cases}}$$

위와같이 입력하여 얻어진 일반방사선사진에서의 판독 결과 얻어진 곡선이 그래프에서 곡선이 된다.

이와 같은 커브를 컬러화한 사진 판독에서 똑같이 시행하여 ROC곡선(윗곡선)을 만든 다음 두 곡선을 비교하여 아래면적이 더 큰 쪽이 병소 판독에 우수하다고 결론 짓는다.

참 고 문 헌

1. Metz CE. Application of ROC analysis in diagnostic image evaluation ; In the physics of medical imaging: Recording system measurements and techniques. New York: Am Assoc Physicists in Med.; 1979 p. 546-72.
2. 久米. ROC曲線: 實驗 畫像評價. p. 65-9.
3. Swets J. ROC analysis applied to the evaluation of medical imaging technique. Invest Radiol 1979; 14: 109-21.
4. Okada R, Shoji S, Horiuchi H. Processing of alveolar bone X-ray Images; Effect of contrast transformation. 日齒保誌 1992; 35: 1026-32.
5. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978; 8: 283-98.
6. Kullendorf B, Grondahl K, Rohlin M, S Henrikson CO. Subtraction radiography for the diagnosis of periapical bone lesions. Endod Dent Traumatol 1988; 4: 253-9.