

# 受试者作业特征曲线(ROC)及其在影像学中的应用

王骏 吴虹桥 宋兆祺

在每天的影像工作中现已有大量的影像设备可以利用,这就提出了为了什么目的去应用什么检测的问题。正因如此,在客观上需要一种方法去比较各种影像诊断的精确性。在过去的20年里,受试者作业特征曲线(ROC)分析法在影像学领域开始应用<sup>[1]</sup>。早在1950年原先使用是为了雷达信号检测的分析,ROC分析法在心理物理学研究中率先应用。1960年Lee Lusted首先认识到ROC分析法在医学判别疾病方面可能会有作用。

敏感性和特异性:需要ROC分析法

传统测试诊断精确的定量方法是敏感性和特异性,这些参数描述病人(有病与没病)的百分数。敏感性或真阳性率(TPF)描述有疾病病人的百分数被正确地判断为阳性结果。特异性或真阴性率(TNF)描述在没有疾病的个体中被判断为阴性结果的机率。敏感性和特异性描述在2分法中判断的结果是:一种判断结果不是阳性就是阴性。

根据它们的性质,许多影像学检查不用2分法判断,而是提供下列3种资料中的一种:①连续定量的资料:病灶的大小用厘米或病灶的CT值用霍斯菲尔德单位,在某些情况下,能显示病灶的病理组织学本质<sup>[2]</sup>,在一个特定的范围内这些数据具有可靠的价值。②标准率的资料:标准率用有限的分类数目在顺序方式中表达一些检测的信息。肾动脉狭窄的程度(狭窄<50%,狭窄50%~74%,狭窄75%~99%和闭塞)能用于进一步诊断和治疗。③定性资料:通常不提供定量资料,许多应用于影像学的标准是形态学性质的,评价病灶的边缘及位置和钙化的出现及性质均能有助于更加明确的诊断。综合这些定性资料可使诊断的2分法得以判定。

当我们将诊断信息转化成是或不是2种回答时,我们需要确定标准或阈值,以告知正常或异常。这种阈值的选择依赖于观察者间和观察者自身的变化。我们能区分过低和过高的判读者,可以看出他们对判定使用不同的阈值。依据临床情况,甚至对同样的影像学检查单独的影像学家将会用不同的阈值。这说明诊断测试的精确度仅用一组敏感性和特异性的值来描述是不合适的。为避免这问题,我们需要不依赖于阈值选择的方法来比较诊断的精确性。

ROC曲线:基本原理

阈值的选择影响敏感性和特异性。对于一个理想的诊断结果的概率分布表明疾病的存在或不存在并不重叠,所选择阈值是在两个分布之间,这种结果的敏感性和特异性都是100%。对于大多数诊断来说,疾病的概率分布和正常分布是重叠的。任何阈值都将导致一些具有疾病的病人错分为正常,或一些没有疾病的个体错分为病人,或两种情况都有。应用低的阈值降

低假阴性结果的数量(高敏感性),但假阳性的数量增加(低特异性);另一方面,增加阈值会增加假阴性(低敏感性),且降低假阳性的数量(高特异性)。这样,在敏感性和特异性之间成互交的关系,一个高的敏感性伴随着低特异性,而一个低的敏感性伴有高特异性。

对所有可能的阈值作计算ROC曲线显示敏感性和特异性之间相互关系。图的纵轴表示敏感性或真阳性率,水平轴表示假阳性率(FPF=1-敏感性)。在ROC曲线上各个作业点表示在给定的一个阈值下敏感性和特异性的组合。在不实际的高阈值下,所有病人都被当作正常分类,导致TPF为0,FPF为0(特异性=1),这与ROC曲线左下角的作业点是一致的。降低阈值既增加TPF又增加FPF(低特异性)。对于可能最低的阈值,TPF和FPF都是1(特异性=0),与ROC曲线右上角相一致。

ROC分析法的实际意义

对于检测诊断利用ROC分析法既提供连续资料又提供等级尺度资料。如果采用5种等级种类对于置信度等级判别通常将产生一个有意义的曲线<sup>[3]</sup>。许多计算机程序能通过观察的作业点计算出一个平滑的ROC曲线。广泛应用的计算机软件包是Metz等开发的<sup>[4]</sup>。这些计算机程序计算出一个副法线ROC曲线。此副法线类型对于ROC分析法通常是被接受的类型并已显示强劲的现实意义<sup>[1]</sup>。在构成ROC曲线实验设计中应避免病例样本的选择偏倚,它有2种来源:①偏倚的范围;②疾病证实的偏倚。既使所有病人利用参考标准评价,如果影像学家进行检测不是盲法则仍然有诊断偏倚的机率。

ROC分析法的应用

1. 比较检测和观察者们:

有病和没病的检测结果,其概率分布重叠的数量决定检测的识别能力,这种重叠决定ROC曲线的形态及位置。如果有病与没病的概率分布是相同的,也即它们完全重叠,TPF和FPF在任何阈值下都相等,这种检测没有识别能力也就没有价值,这种检测的ROC曲线从图的左下角到右上角是直对角线,此“曲线”下的区域是0.5(整体区域的50%)。在另一方面,一个理想的检测在分布上没有重叠,ROC曲线有最佳作业点(即TPF=1和FPF=0),相当于ROC图的左上角,在这ROC曲线下的区域为1.0(全部区域为100%)。

ROC曲线下的区域是检测的诊断精确性的量度,常用于诊断检测之间和观察者们之间的比较<sup>[1]</sup>。运用适当的计算机软件,能够计算出ROC曲线下的区域并对显著性差异用单一的Z-score检验作检测<sup>[4]</sup>。但曲线下区域的非参数计算,对于比较ROC曲线下的区域非参数方法比Z-score检验更合适。因此,ROC曲线下区域差异的意义可随分析方法的改变而改变。

依据ROC曲线下区域的比较检测的主要优势是不依赖于

诊断标准, 这样在敏感性和特异性评价上消除了阈值的影响。但这样做会出现另一个问题, ROC 曲线的大部分是由临床不相关的 TPF 和 EPF 的组成。ROC 曲线的末端角落表示高的敏感性和低的特异性组合或反之亦然。在许多临床资料中这些组合是没用的。如果 2 种不同检测的 ROC 曲线相交此问题尤为相关。无论何时临床上要求高的敏感性, 我们要应用检测高的敏感性范围内一个高的 ROC 曲线。然而, 在 ROC 曲线相交时, 如比较检测的诊断精确性, 则曲线下区域的用处有限<sup>[5]</sup>。为了这个问题, 对于 ROC 曲线的局部评价提出了 2 种方法。McGlish 主张依据 FPF 值的一定范围作一部分 ROC 曲线的分析<sup>[6]</sup>, 此法允许在预置的特异性区域内对检测的诊断能力作出比较。Jiang 等提出部分区域指数 (PAI), 采用预置敏感性的区域<sup>[7]</sup>。PAI 的计算: 感兴趣的 ROC 曲线部分之下的估算区域被此部曲线下的最大可能区域除 (TPF 值的范围在预定的 TPF0 之上)。用这种方法产生了指数值, 范围从 0 到 1, 可用来作比较类似曲线下的全部区域。

### 2. 最佳的阈值:

ROC 曲线的另一个潜在的作用是检测的最佳阈值。ROC 曲线包含在所有可能的阈值上所有敏感性和特异性的组合。这为临床实践提供了估价最佳阈值的机会。概括地说, ROC 分析法对于比较影像学检测和观察者的诊断精确性是一种有用的技术。由于阈值的影响被排除, 曲线下的区域提供一种检测诊断精确性的客观参数, 优于单纯比较敏感性和特异性。因为只有 ROC 曲线的一部分代表临床相关的敏感性和特异性组合, 在相关的敏感性和特异性范围中比较 ROC 曲线比比较曲线下的全部区域还好。

### 3. 应用:

①监视器 (CRT)-1 采用传统模拟的、高分辨力的视频监视器; 监视器 2 为影像数据的数字传递<sup>[8]</sup>。作者用 ROC 分析法, 采用模拟人胸部的体模对其进行评价。对每种 CRT 的 100 幅

影像由 5 位观察者参与评定。从相同设备 100 个不同的体模曝光中产生这些影像。从单个专家的数据和全体专家们的数据依据标准步骤计算出 ROC 的点。在正常概率图中每个 ROC 曲线下区域, 参数 AZ, 由 Swets 和 Pickett 修正的 Dorfman 和 Alf 计算程序, 给出由最小二乘拟合合法得到的点对最大似然估计量 AZ 的 ROC 点的 ROC 曲线。用 *t* 检验计算其显著性。结果: 左上肺小型不规则球形病灶 (0.8: 0.78), 上纵隔大型球形病灶 (0.83: 0.82), 右上肺气胸结构 (0.91: 0.90), 两者间在统计学上无显著性差异。结论: 在 ROC 试验中用拟人胸部体模, 2 种 CRT 在仿真病理学的检验中没有显著性差异。

②在临床实践中, 判读胸部 X 线照片大约 80% 与先前的 X 线检查相对照, 对照判读可帮助影像学家识别异常。在数字 X 线照片的情况下, 能从当前的 X 线照片减去先前的 X 线照片产生即时减影图像以增强间隔性改变的区域。为此, 作者对即时减影图像进行了客观评价<sup>[9]</sup>。观察者测试了 50 个病例, 11 个放射学家充当观察者, 对当前和先前影像以及即时减影图像组成比较。用最大似然估计量的方法使副法线 ROC 适合于各个观察者的置信度等级。当它在单位方格纸上被描绘时, 在副法线 ROC 曲线下的区域代表指数 AZ, 并对每个相应的曲线作计算。对一组数据应用曲线二端 *T* 检验于阅读者各自的 AZ 值, 证明用和不用即时减影所获取的 ROC 曲线在统计学上具有显著差异。为了表达一组观察者总体诊断精确, 通过不同判读者各自曲线的平均斜率计算复合 ROC 曲线。结果: 和即时减影图像一起观察成对的先前和当前数字胸部 X 线照片时, 单个和复合 ROC 曲线证实对提高诊断的精确度具有显著改善。判读者各自 ROC 曲线下区域的 AZ 值, 用即时减影平均 AZ 值从 0.89 增加接近 0.98。结论: 观察成对数字化先前和先前胸部 X 线照片以及即时减影图像时, 数字减影技术对于胸部 X 线照片能改善诊断的精度。

(1999-05-31 收稿)

## 综览 一外伤后脊髓空洞症

很多严重脊髓外伤患者得以长期存活, 然而可能出现不少并发症, 这包括外伤后空洞、微小囊性变性或蛛网膜囊肿形成。脊髓空洞的平均形成时间是 5~8 年 (2 个月~32 年)。随着 MRI 使用的增多, 脊髓空洞的诊断也渐增多, 1976 年在截瘫和四瘫患者为 2.3%, 1985 年为 3.2%。在另一组患者自 1991 年至 1993 年增加了 50%; 1995 年有报告其发病率为 4.45%。作者医院最近收治了 440 例脊髓空洞症, 其中 140 例是外伤后引起。形成脊髓空洞的因素包括出血、缺血、脊髓周围通路受阻和局部脊髓纤维化。临床表现主要是感觉障碍、运动功能丧失、疼痛和深肌腱反射变化。MRI 所见为外伤上下的脊髓内纵行囊性空洞, T<sub>1</sub>WI 低信号而 T<sub>2</sub>WI 高信号。为了治疗设计 MRI 应注意包括其上端和下端。

## 心肌梗死—MRI 的作用?

MRI 近年来对心脏的诊断有较大进展。它将使冠心病和心肌梗死的无创性评价成为现实。本文对评价心梗后的心肌解剖、功能、灌注和冠脉解剖的现状和潜力作一综览。电影 MRI 对局灶和全面的心壁运动障碍提供高时间/空间分辨率的可靠分析。标记 (Tagging) 技术 (译者注: 图像上显示清晰或模糊的网格) 能将部分心肌作标记并识别其三维运动模式。MRI 还可测定心肌灌注以及药物负荷下心肌运动的障碍。独特的是, 心肌水肿藉良好的软组织对比而得以显示。对目前还受限的冠脉及其分流手术的评价进行了讨论。

郭俊渊摘译自 DE Kivelitz. R<sup>+</sup>Fo 171, (1999), 349-358

## • 外刊摘要 •