文章编号:1000-5404(2008)20-1889-04

论著

基于Bayes理论的计算机辅助诊断系统在孤立性肺结节CT诊断中的应用

陈 伟,刘进康,李文政,熊 曾,龙学颖,周 晖 (中南大学湘雅医院放射科,长沙 410008)

摘 要:目的 初步探讨基于Bayes理论的计算机辅助诊断(computer-aided diagnosis, CAD)系统在孤立性肺结节(solitary pulmonary nodule, SPN)CT诊断中的价值。**方法** 依据Bayes理论先从352例SPN训练集(恶性135例,良性217例)中求出恶性SPN的验前比及各临床和CT表现的似然比,再运用VC^{++}语言编制基于Bayes理论的CAD系统,用它计算每个SPN的恶性概率,并前瞻性地检验该系统在132例SPN测试集(恶性61例,良性71例)中的诊断效能,与2位高年资和2位低年资放射科医师常规阅片的表现作比较。**结果** 成功构建基于Bayes理论的CAD系统,它诊断训练集SPN的敏感度、特异度和符合率分别为88.9%、93.1%、91.5%,诊断测试集SPN的敏感度、特异度、符合率、阳性预测值及阴性预测值分别为88.5%、85.9%、87.1%、84.4%、89.7%,其诊断符合率与高年资甲、乙医师比较无统计学差异($P>0.05$),但高于低年资丙、丁医师($P<0.05$)。**结论** 基于Bayes理论的CAD系统可帮助医师尤其是低年资医师提高鉴别SPN良恶性质的能力,并在指导SPN的临床决策中有一定的参考作用。

关键词: 硬币病变,肺;体层摄影术,X线计算机;诊断,计算机辅助;Bayes理论
中图法分类号: R319;R563.04;R814.42 **文献标识码:** A

Application of Bayesian theory based computer-aided diagnosis system in CT diagnosis of solitary pulmonary nodules

CHEN Wei, LIU Jin-kang, LI Wen-zheng, XIONG Zeng, LONG Xue-ying, ZHOU Hui (Department of Radiology, Xiangya Hospital, Central South University, Changsha 410008, China)

Abstract: Objective To preliminarily explore the value of computer-aided diagnosis (CAD) system based on Bayesian theory in the diagnosis of solitary pulmonary nodules (SPNs) with CT. **Methods** Totally 352 consecutive SPN cases (malignancy $n=135$, benignity $n=217$) were collected retrospectively to form the training set. According to Bayesian theory, the prior odds of malignant SPNs and the likelihood ratios of clinical and CT findings were derived from the training set, then these derived values were used to construct a Bayesian theory based CAD system by using VC^{++} language. Utilizing this system, the probability of malignancy in each case with SPN was calculated, and SPNs with $\geq 50\%$ calculated probability were judged as malignancy and those with $<50\%$ calculated probability was judged as benignity. This system was also tested prospectively for its diagnostic validation on the test set (malignancy $n=61$, benignity $n=71$), compared with the performance of the two chest radiologists and two radiologic residents using routine diagnostic method. **Results** The Bayesian theory based CAD system was constructed successfully. The sensitivity, specificity, accuracy of this system for the training set were 88.9%, 93.1%, 91.5%, respectively. On the test set, the sensitivity, specificity, accuracy, positive predictive value, negative predictive value of this system were 88.5%, 85.9%, 87.1%, 84.4% and 89.7%, respectively. The accuracy of this system showed no statistical significance with that made by chest radiologist A (80.3%, $\chi^2=2.37$, $P=0.122$) and B (79.5%, $\chi^2=3.12$, $P=0.076$), but was higher than that by radiological residents C (74.2%, $\chi^2=7.05$, $P=0.012$) and D (74.2%, $\chi^2=6.56$, $P=0.009$). **Conclusion** The Bayesian theory based CAD system can help physicians differentiate benign

基金项目: 湖南省自然科学基金(07JJ5010)
Supported by the Natural Science Foundation of Hunan Province (07JJ5010)
作者简介: 陈 伟,男,湖南省祁东县人,硕士研究生,主要从事CT图像后处理及质量控制方面的研究。电话:(0731)4327441,E-mail: chenwei_rt@sina.com
通信作者: 刘进康,电话:(0731)4327441,E-mail: liujinkangdoctor@sina.com
收稿日期: 2008-07-15;**修回日期:** 2008-08-12

from malignant SPNs, especially for less experienced physicians. This system can provide a reference in determining the management of SPNs.

Key words: coin lesion, pulmonary; tomography, X-ray computed; diagnosis, computer-assisted; Bayes theorem

孤立性肺结节(solitary pulmonary nodule, SPN)是指在X线上表现为肺内单发的直径≤3 cm的圆形或类圆形不透光区,不伴有区域性淋巴结肿大、阻塞性肺不张或肺炎[1]。SPN的定性诊断往往是放射和临床医生共同面临的一个挑战。尽管近年以研究分类决策树、神经网络等高级数据挖掘技术辅助诊断SPN为热点,但是传统的计量诊断方法因其可以预测SPN的恶性概率而具有较大的临床应用价值[2]。本研究运用计量诊断中的Bayes分析结合计算机技术构建了1个简便易行的计算机辅助诊断(computer-aided diagnosis, CAD)系统,以期为SPN的定性诊断和临床决策提供帮助。

1 资料与方法

1.1 临床资料

病例纳入标准:①CT上表现为肺内单发的长径与短径平均值≤3 cm的圆形或类圆形不透光区;②可出现空洞、钙化或伴有卫星病灶;③经病理或临床证实者。排除标准:①伴有区域性淋巴结肿大、阻塞性肺不张或肺炎;②图像存在明显伪影致形态学评价或CT值测量不可靠者。搜集我院(三级综合性医院)CT室2005年1月至12月的连续SPN病例352例作为本研究的训练集,其中,恶性135例,男性92例,女性43例,年龄19~87岁,平均59.0岁;良性217例,男性149例,女性68例,年龄7~81岁,平均44.4岁。另搜集我院2007年3月至6月经CT发现的132例SPN作为测试集,其中,恶性61例,男性44例,女性17例,年龄19~76岁,平均57.1岁;良性SPN 71例,男47例,女24例,年龄19~70岁,平均47.7岁。全部病例中,196例恶性SPN均经手术或穿刺活检的病理证实,包括鳞癌26例,腺癌82例,细支气管肺泡癌28例,孤立性转移瘤15例,组织类型未明的恶性肿瘤45例;除40例结核瘤和21例炎性假瘤经病理证实外,其余227例良性SPN由CT随访缩小、消失及2年以上稳定等临床途径证实。除测试集中56例经临床证实者为随访截尾时病例,其余均为新发病例。追查每例SPN患者的临床信息,包括年龄、发病时的主要症状、吸烟史及有无恶性病史等。

1.2 CT扫描方法

采用PHILIPS Brilliance 16层螺旋CT机。扫描范围自肾上腺水平至肺尖,平静呼吸下屏气时扫描。无碘对比剂禁忌证,平扫图像无较大空洞和弥漫性钙化,直径10~30 mm的SPN患者行增强扫描。增强扫描使用非离子型对比剂和高压注射器,注射总量按每千克体质量1.5 ml计算,注射流率为2 ml/s,常规增强扫描的延时时间为80 s,动态增强方案为注药开始后第1、2、3、4、5分钟对全结节行螺旋扫描。主要技术参数为:120 kVp,100~250 mAs,探测器组合16 mm×1.5 mm,螺距值0.6~0.9,重建层厚2 mm,重建间隔1 mm,重建矩阵512×512。

1.3 图像测量及评价

CT图像的测量和征象评判均在工作站中的实时MPR图像上进行。结节大小测量的窗宽为1 200 Hu,窗位为-600 Hu。强化值测量的窗宽为200 HU,窗位为40 HU,兴趣区选择结节强化最明显的实质区,取增强后CT值(动态增强为强化峰值)与平扫CT之差值作为结节的强化值。结节的形态学评判参照国内外文献公认的标准,部分可量化指标规定如下:结节出现1个以上弧弦距与弦长比值≥4/10的分叶征象归为"深分叶"(包括棘突征),余为"浅分叶";结节有≥2条较平直或成角的规整边缘为"多边形"的轮廓形态[3];以长度>5 mm毛刺为主的毛刺征归为"长毛刺",余为"短毛刺";结节内出现直径>5 mm的透光区为空洞,而仅出现直径≤5 mm的透光区为空泡征。先由2名具有15年以上胸部CT阅片经验的高年资放射科医师甲、乙盲法分别评价全部病例的SPN形态学征象,存在分歧时通过协商达成一致,并独立估计测试集SPN的恶性概率。再由具有5年左右工作经验的低年资放射科住院医生丙、丁参照高年资医师已认定的CT征象,各自独立作出测试集SPN恶性概率的估计。

1.4 Bayes理论及CAD系统的建立

本研究运用Bayes理论中的似然比和相对比(odds)形式[4],将恶性SPN的验前比、验后比及对定性诊断可能有意义的临床表现和CT征象以线性方式关联起来,并以此计算SPN的恶性发病概率。用下列公式表示:

$$O_{pre}=\frac{O_{pre}}{1-O_{pre}} \quad \text{公式(1)}$$

$$LR_M=\frac{P(X_a/Y_m)}{P(X_a/Y_b)} \quad \text{公式(2)}$$

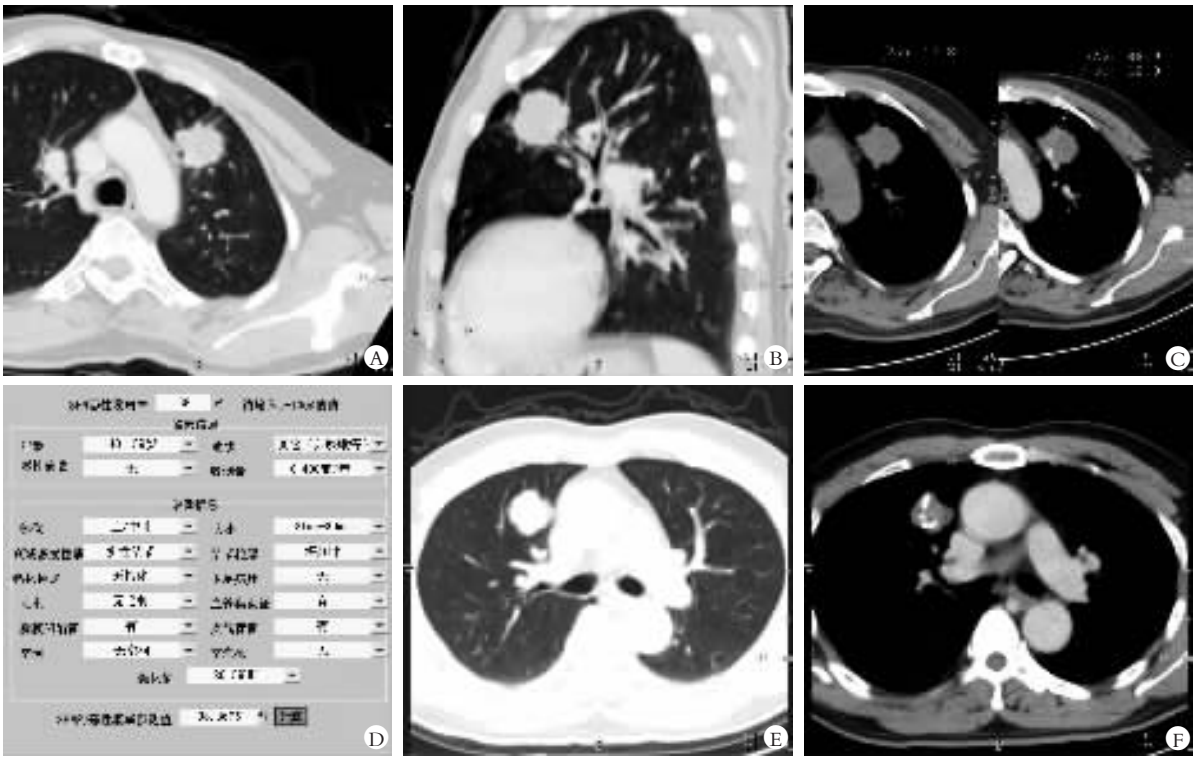
$$O_{pos}=O_{pre}\cdot LR_1\cdot LR_2\cdots LR_m \quad \text{公式(3)}$$

$$O_{pos}=\frac{O_{pos}}{1+O_{pos}} \quad \text{公式(4)}$$

其中P_{pre}、P_{pos}分别代表恶性SPN的验前概率和验后概率,O_{pre}、O_{pos}分别代表恶性SPN的验前比和验后比,$P(X_a/Y_M)$、$P(X_a/Y_B)$分别代表某判别指标X_a在恶性SPN(Y_M)和良性SPN(Y_B)中的发生概率(可由样本频率估计),LR代表某判别指标的似然比(likelihood ratio, LR),用来表示某临床表现或影像学特征出现或不出现时SPN恶性发病与良性发病机会的比例,其取值0至无穷大,越大于1,越提示恶性诊断;越趋于0,越提示良性诊断;而取值为1时,对SPN的良恶性诊断无任何鉴别意义。本研究参照文献纳入4项临床指标和13项CT指标作为鉴别SPN良恶性质的判别指标,共47种状态(表1),将年龄、

表 1　SPN 相关临床信息和 CT 表现在不同状态下的似然比

项目	似然比	项目	似然比	项目	似然比	项目	似然比
(1)年龄		(5)部位		(9)钙化模式		(10)支气管征	
<40 岁	0.19	上/中叶	1.08	无钙化	1.33	有	1.85
40 ~ 59 岁	1.06	下叶	0.89	良性钙化	0.03	无	0.72
≥60 岁	2.79	(6)结节大小		难定性钙化	0.80	(15)空洞	
(2)症状		4 ~ 10 mm	0.24	(10)卫星病灶		无空洞	1.12
基本无症状	1.15	11 ~ 20 mm	0.77	有	0.19	良性空洞	0.20
以咯血为主	1.24	21 ~ 30 mm	1.66	无	1.43	恶性空洞	1.61
以胸痛为主	0.95	(7)衰减密度性质		(11)毛刺征		(16)空泡征	
以发热为主	0.23	实性结节	0.84	无毛刺	0.74	有	4.98
其他(咳嗽等)	0.88	部分实性结节	2.61	长毛刺	0.63	无	0.81
(3)吸烟量		非实性结节	3.21	短毛刺	4.57	(17)强化值	
<400 支/年	0.70	(8)结节轮廓		(12)胸膜凹陷征		<20 Hu	0.08
400 ~ 799 支/年	2.25	平滑	0.26	有	1.78	20 ~ 59 Hu	2.98
≥800 支/年	3.43	浅分叶	1.16	无	0.41	≥60 Hu	0.19
(4)恶性病史		深分叶	3.88	(13)血管集束征			
有	3.42	多边形	0.14	有	2.26		
无	0.91			无	0.17		



A ~ D 为同一 SPN 病例,手术病理为低分化腺癌,男性,41 岁,咳嗽、咳痰,无恶性病史,不吸烟。A:示左肺上叶大小为 29 mm × 27 mm × 30 mm 的实性结节,见深分叶和胸膜凹陷征;B:示支气管征和血管集束征;C:示强化值为 41 Hu;D:示基于 Bayes 理论的 CAD 系统预测该 SPN 的恶性概率为 98.4%。E 和 F 为同一 SPN 病例,手术病理为错构瘤,男性,56 岁,体检时胸片发现异常,无恶性病史,吸烟 36 年,30 支/日,基于 Bayes 理论的 CAD 系统预测其恶性概率为 0.6%。E:示右肺上叶大小为 27 mm × 25 mm × 26 mm 的实性结节,见浅分叶;F:示良性钙化

图 1　基于 Bayes 理论的 CAD 系统诊断 SPN 实例

吸烟量、结节大小和强化值等计量资料转换成等级资料,并假定各判别指标和指标下的各状态相互条件独立。在 352 例训练集中,运用公式(1)、公式(2)分别求出恶性 SPN 的验前比和各判别指标在不同状态下的似然比。再根据所得验前比、似然比及公式(3)和公式(4),运用 VC^{++} 语言在 Microsoft Visual C^{++} 6.0 编译环境下制作了一个基于 Bayes 理论的 CAD 系统,运用它计算全部 SPN 的恶性概率,计算概率≥50% 判为恶性,<50% 判为良性。

2　结果

恶性 SPN 的验前比为 0.61,SPN 相关的各项临床信息和 CT 表现在不同状态下的似然比见表 1。

根据表 1 结果和 Bayes 分析的公式,成功构建基于 Bayes 理论的 CAD 系统(图 1)。该系统回代检验 352 例训练集 SPN,

其敏感度、特异度和符合率分别为88.9%、93.1%、91.5%，它前瞻性地检验132例测试集SPN的阳性预测值和阴性预测值分别为84.4%、89.7%，与4位医师常规阅片法诊断SPN的真实性指标及其比较见表2。

表2 基于Bayes理论的CAD系统与4位医师常规阅片法诊断132例测试样本的真实性指标及其比较

诊断方式	敏感度（%）	特异度（%）	符合率（%）	χ^2值	P值
CAD系统	88.5	85.9	87.1	-	-
高年资甲医师	77.0	83.1	80.3	2.37[a]	0.124[a]
高年资乙医师	86.9	73.2	79.5	3.12[a]	0.077[a]
低年资丙医师	55.7	90.1	74.2	7.05[a]	0.008[a]
低年资丁医师	72.1	76.1	74.2	6.56[a]	0.010[a]

a代表4位医师常规阅片法与基于Bayes理论的CAD系统诊断符合率的比较；“-”为无数据

3 讨论

SPN是胸部放射线检查常见的病变之一，大多在胸部X线片上被偶然发现，近年随着肺癌CT筛查的应用和多层螺旋CT的普及，被发现的SPN越来越多，其诊断和处理一直是临床上的难点。

SPN的病因多达二三十种，引人关注的焦点还是在SPN良、恶性的二分类定性诊断，旨在尽可能切除可治愈的恶性结节和避免良性结节的不必要手术。但是，与SPN相关的临床信息和影像学表现在良恶性SPN中存在部分重叠，导致多数SPN的性质难以确定或诊断正确率不高，尤其体现在经验欠缺的低年资医生的诊断过程中。Bayes理论指出：如果事件的结果不确定，那么量化它的唯一方法就是该事件的发生概率。本研究运用的CAD系统能够综合SPN的恶性发病率、SPN相关的各临床及CT表现等已知信息计算预报SPN恶性的发病概率。从表2得知，运用该系统预报的恶性概率对SPN进行定性判断的符合率与经验丰富的高年资医师对SPN定性诊断的符合率相当（$P>0.05$），明显高于经验欠缺的低年资医师（$P<0.05$）。此外，运用该系统预报的恶性概率来指导SPN处理措施的抉择具有一定的实际意义。譬如，参照相关权威组织提供给SPN患者的建议[5,6]，低的恶性概率采取CT随访等温和的观察手段，高的恶性概率可以立即进行手术，而中等的恶性概率可以行PET检查或穿刺活检等，如此以较低医疗成本和风险获得最佳效果。

数据挖掘是一种能从大量的、模糊的随机数据中提取有用知识的信息处理技术。它用在SPN的诊断中，极大地扩大了医生有限的个人知识和经验，可以提高分析判断的客观性和全面性，且可重复性好，不易疲劳。在CAD系统中用到的数据挖掘方法主要有统计方法（计量诊断）、决策树、神经网络，以及遗传算法、模糊逻辑等技术。与同为计量诊断的最大似然法相比，本系统利用了SPN的验前概率，判别的符合率有所提高；与Logistic回归严格的判别指标筛选相比，本系统充分利用了SPN相关的临床和影像信息，符合医生常规的诊断思维；与人工神经网络“暗箱式”推理的不可理解性和复杂性相比，本系统具有概率推理、逻辑简单、可处理缺省数据和操作性好的特点；与分类决策树赋值的无特定意义相比，本系统报告的恶性概率可用来定性SPN和指导SPN的临床决策[2]。

尽管，基于Bayes理论的CAD系统有相当于高年资医生的诊断符合率，但对于个体患者来说，运用该方法仍然可能在其身上发生不可料想的误诊或漏诊。因此，该系统可以作为医师在鉴别SPN时的一个有效辅助工具，但并不能取代医师的地位作完全的“自动化”诊断。

参考文献：

[1] Jeong Y J, Yi C A, Lee K S. Solitary pulmonary nodules: detection, characterization, and guidance for further diagnostic workup and treatment[J]. AJR Am J Roentgenol, 2007, 188(1): 57-68.

[2] 陈 卉, 王晓华. 数据挖掘技术在计算机辅助肺癌诊断中的应用[J]. 中国组织工程研究与临床康复, 2007, 11(5): 879-881, 885.

[3] Furuya K, Murayama S, Soeda H, *et al*. New classification of small pulmonary nodules by margin characteristics on high-resolution CT[J]. Acta Radiol, 1999, 40(5): 496-504.

[4] Gurney J W. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part I. Theory[J]. Radiology, 1993, 186(2): 405-413.

[5] Khan A. ACR Appropriateness Criteria on solitary pulmonary nodule [J]. J Am Coll Radiol, 2007, 4(3): 152-155.

[6] Gould M K, Fletcher J, Iannettoni M D, *et al*. Evaluation of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition) [J]. Chest, 2007, 132(3 Suppl): 108S-130S.

（编辑 王 红）