

基于CT图像特征的肺腺癌预后因素分析

鲁晓腾, 龚敬, 聂生东

上海理工大学医学影像工程研究所, 上海 200082

【摘要】目的:提出一种基于CT图像特征的肺腺癌预后因素分析方法,旨在探究不同种类CT图像特征对肺腺癌预后的影响。**方法:**首先,对肺部肿瘤进行分割和特征提取;然后,使用Kaplan-Meier方法进行单因素生存分析;使用COX回归模型进行多因素生存分析,得到肺腺癌的独立预后因素。最后,利用支持向量机(SVM)建立分类器对独立预后因素的预后能力进行检验。**结果:**选用Lung CT-Diagnosis数据库中61例患者进行试验,单因素分析显示径向方差、边缘粗糙度、GLCM熵以及GLCM非相似性与肺腺癌患者生存率显著相关($P<0.05$)。COX回归模型多因素分析发现唯有径向方差与肺腺癌患者生存显著相关($P<0.05$)。SVM分类器分类结果显示径向方差能够在一定程度上对患者生存时间进行预测。**结论:**通过对比分析,发现径向方差、边缘粗糙度、GLCM熵、GLCM非相似性与肺腺癌预后有关;径向方差是肺腺癌的独立预后因素。通过提取分析上述图像特征,医生可以对肺腺癌患者进行更加精准的预后进而延长患者生存时间。

【关键词】肺腺癌;预后;图像特征;独立预后因素

【中图分类号】R318;R734.2

【文献标志码】A

【文章编号】1005-202X(2019)03-0291-05

CT image feature-based analysis on the prognostic factors of lung adenocarcinoma

LU Xiaoteng, GONG Jing, NIE Shengdong

Institute of Medical Imaging Engineering, University of Shanghai for Science and Technology, Shanghai 200082, China

Abstract: Objective To propose a method for the analysis of the prognostic factors of lung adenocarcinoma based on CT image features, and explore the effects of different kinds of CT image features on the prognosis of lung adenocarcinoma. **Methods** Firstly, the lung tumors were segmented and their features were extracted. Secondly, Kaplan-Meier method was used to perform univariate survival analysis, and a multivariate survival analysis was carried out with COX regression model to obtain independent prognostic factors. Finally, a classifier based on support vector machine was established to test the prognostic ability of independent prognosis factors. **Results** The data of 61 patients with lung adenocarcinoma were selected from Lung CT-Diagnosis dataset. The univariate analysis showed that several image features, including radial variance, edge roughness, GLCM entropy and GLCM non-similarity, had significant effects on the overall survival of patients with lung adenocarcinoma ($P<0.05$). The multivariate analysis based on COX regression model revealed that only radial variance was significantly associated with the survival of patients with lung adenocarcinoma ($P<0.05$). The result of classifier based on support vector machine showed that to some extent, using radial variance could predict the survival time of patients. **Conclusion** Four features, namely radial variance, edge roughness, GLCM entropy and GLCM non-similarity, are proved to be associated with the prognosis of patients with lung adenocarcinoma. Moreover, radial variance is regarded as the independent prognostic factor of lung adenocarcinoma. The analysis of the above image features can provide doctors with more accurate prognosis which is helpful for prolonging the survival time of patients with lung adenocarcinoma.

Keywords: lung adenocarcinoma; prognosis; image feature; independent prognostic factor

【收稿日期】2018-09-06

【基金项目】国家自然科学基金(60972122);上海市自然科学基金(14ZR1427900)

【作者简介】鲁晓腾,在读硕士,研究方向:医学图像处理,E-mail: lxt0420wz@163.com;龚敬,在读博士,研究方向:医学图像处理,E-mail: gongjing1990@163.com

【通信作者】聂生东,博士,教授,研究方向:医学图像处理,E-mail: nsd4647@163.com

前言

在过去50年,肺癌是全球范围内发病率和死亡率增长最快的恶性肿瘤,稳居我国恶性肿瘤之首。其中,非小细胞肺癌(Non-Small Cell Lung Cancer, NSCLC)占肺癌患病总人数的80%~85%^[1]。作为最常见的肺癌组织学类型之一,肺腺癌占NSCLC总数

的50%以上^[2]。随着科技的进步和医疗的发展,肺腺癌患者的存活状况得到了一定的改善。但肺腺癌起病较为隐匿,当下面临的仍然是5年生存率在15%左右这一严峻事实^[3]。因此,对肺腺癌的预后因素进行研究对于改善患者的生存状况有重要意义。

针对肺腺癌预后因素的研究,已经存在一些相关工作。Grove等^[4]发现凸度和熵率两个因素是肺腺癌的独立预后因素。Hawkins等^[5]提取一系列图像的三维特征来分析探究肺腺癌的独立预后因素,并使用决策树方法建立分类器对结果进行检验。Balagurunathan等^[6]对多种二维和三维特征进行提取分析,发现病程长度灰度级不均匀度有较好的预后能力。杨为贵等^[7]回顾性研究了109例肺腺癌患者的临床资料,采用Kaplan-Meier法和多因素回归分析法对临床因素进行生存分析。结果显示,肿瘤大小、TNM分期等因素是肺腺癌的独立预后因素。廉政君等^[8]回顾性分析了201例晚期非小细胞肺癌患者的临床资料。结果发现:患者吸烟情况以及治疗方案选择是影响患者生存的独立预后因素。

从目前的研究成果来看,肺腺癌预后方面的研究普遍存在两个问题。其一是提取的预选因素没有完全包括直方图统计特征、形状特征以及纹理特征3类极具代表性的图像特征。例如文献[4]中只提取了部分形状特征和部分纹理特征;文献[5]只提取了部分三维形状特征和部分纹理特征。其二是对于生存分析获得的独立预后因素没有通过实验加以验证。例如文献[7-8]只进行了预后因素的筛选,并没有对结果进行实验验证。

针对目前研究存在的不足,本文设计了新的预后研究方法,基于多类特征对肺腺癌的预后因素进行分析探究;同时,设计分类器检测独立预后因素的预后能力。进而完善目前研究中存在的不足,获得更好的预后效果。

1 材料和方法

1.1 材料

本研究所用数据来自于TCIA(The Cancer Imaging Archive)公共访问中的Lung CT-Diagnosis数据库^[9]。该数据库中共有61组肺腺癌患者的CT序列图像,每组序列平均包含75幅图像,每一幅图像的大小为512像素×512像素,图像层厚在2.5~6.0 mm之间。实验平台为64位Windows 10操作系统,i7-4770-3.4 GHz处理器,8 GB内存;使用的软件是Matlab 2015a和SPSS 22.0。

1.2 方法

首先,进行肺实质和肿瘤区域的提取。接着,完成图像部分的特征提取。然后,使用SPSS软件绘制生存曲线图以及单因素分析和多因素分析。最后,使用支持向量机(Support Vector Machine, SVM)建立分类器对独立预后因素的预后能力进行验证。实验方法流程如图1所示。

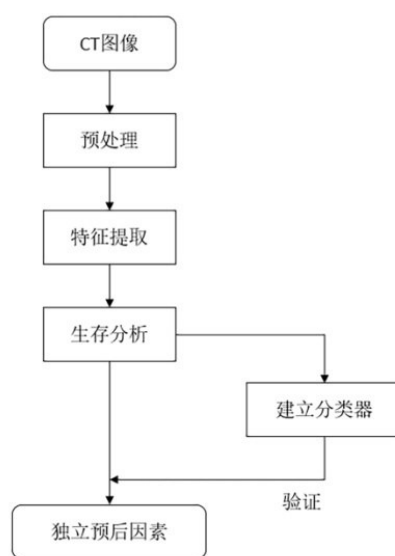


图1 方法流程图

Fig.1 Flow chart of experimental method

1.2.1 预处理 为了便于感兴趣区域(Region of Interest, ROI)的提取,需先将肺实质部分提取。本文采用Ostu阈值法与数学形态学相结合的方法对肺实质进行提取。提取肺实质后,为避免分割方法对预后造成影响,本文采用稳定性较高的区域生长法和基于边缘检测的方法对ROI进行分割。

1.2.2 特征提取 直方图统计特征是根据ROI的灰度分布直方图求取出的一系列特征。有研究称:直方图统计特征可以在肺结节良恶性诊断方面作为重要的依据^[10];且与肺腺癌预后有紧密的联系。这说明该类特征有极大可能性与肺腺癌预后相关。形状特征是一类较为直观且容易理解的图像特征。临床医学上,许多医生就是根据肿瘤的分叶征、毛刺征等形状特征对肿瘤的良恶性以及患者生存信息进行判断。纹理特征包含了物体表面的性质和结构,还在一定程度上反映了物质与周围环境的关系^[11]。几年来,很多研究表明:纹理特征与肺癌的肿瘤分期、转移、存活情况相关^[12-14]。所以,本文提取了灰度共生矩阵(Gray Level Co-occurrence Matrix, GLCM)纹理特征作为备选特征。以上特征的提取方法在以往文献中有所介绍,本文不再赘述。本实验提取各类特征共计30种,具体特征如表1所示。

表1 实验中提取的特征

Tab.1 Features extracted in experiment

特征类型	特征名称
直方图特征	均值、方差、偏度、峰态、能量、熵
形态学特征	面积、周长、径向均值、径向方差、边缘粗糙度、体积等
纹理特征	GLCM非相似性、GLCM熵、GLCM差、GLCM逆差矩、GLCM角二阶矩等

1.2.3 生存分析 生存分析是一种将终点事件和此事件发生所经历的时间结合分析的一种统计分析方法。本文主要使用如下3种方法进行生存分析:

(1)生命表分析。生命表法分析是通过计算落入时间区间 $[tk-1, tk]$ 内的失效和删失的观察个数来估计该区间上的死亡概率,然后用该区间及其之前各区间上的生存概率之积来估计假设生存率^[15]。

(2)Kaplan-Meier分析。Kaplan-Meier分析法又称乘积极限法,是在1985年由Kaplan和Meier提出的。它根据每一个事件发生时间点的条件概率的估计和事件相应的概率等信息来估计每一个时间点的生存率^[16]。

(3)COX回归分析。COX回归分析是一种存在删失数据情况下拟合时间—时间模型的一种方法^[17]。COX回归分析需要建立生存时间随危险因素变化的回归模型,进而确定对生存时间有影响的独立预后因素。

在对实验数据进行分析后,本文使用生命表法绘制生存曲线图,使用Kaplan-Meier法对图像特征进行单因素分析,使用COX回归分析法对预后因素进行多因素分析。

1.2.4 建立分类模型 SVM是一种基于统计学习理论的机器学习方法。相较于其他机器学习方法,SVM有如下优势:首先,由有限的训练集样本得到的小误差仍可以保证对独立测试集保持小的误差;其次,支持向量算法是一种凸优化算法,局部最优解一定是全局最优解;另外,SVM是一种结构风险最小化算法,泛化能力较强。因此,本文使用SVM建立分类器对肺腺癌的独立预后因素的预后能力进行检验。

为充分利用有限数据,本文使用留一交叉验证法对数据进行处理。这种方法计算相对繁琐,但数据利用率更高,得到的结论也更准确,更适合小样本数据的分类检测。

2 结果

2.1 总体生存曲线分析

本次试验,参与随访的患者一共有61例。到随

访日期截止时,仍有38例存活。患者不足3年的存活率是60.7%,3年存活率为39.3%,5年存活率为1.6%。在SPSS软件中,使用生命表分析方法,将患者的生存时间,存活状态作为输入变量,可以生成患者的生存曲线图,如图2所示。

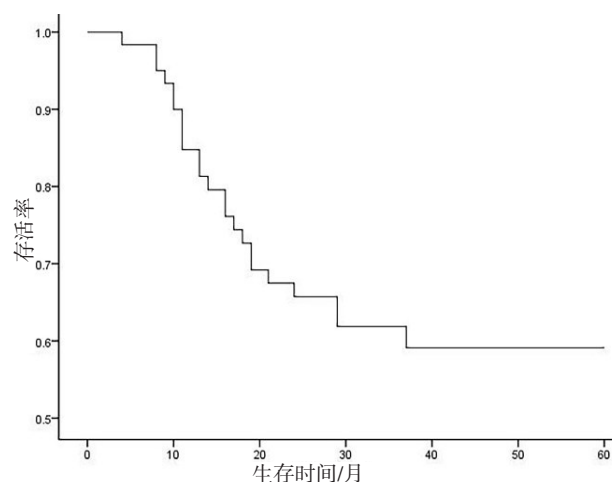


图2 生存曲线图

Fig.2 Survival curve

2.2 单因素分析

求取各组特征数据的中位数作为其cut-off值,然后将患者的生存时间、存活状态以及特征数据依次输入到Kaplan-Meier分析模型中进行单因素分析。为避免ROI的分割方法对结果造成影响,本文对两种分割方法的分割结果及二者平均后的数据分别进行了单因素分析。

对区域生长法的ROI分割结果进行单因素分析发现:灰度均值、径向方差、边缘粗糙度、GLCM非相似性和GLCM熵与患者的生存显著相关($P<0.05$)。具体数据见表2。

对边缘检测法的ROI分割结果进行单因素分析发现:径向方差、边缘粗糙度、GLCM角二阶矩、GLCM差熵、GLCM非相似性、GLCM熵和GLCM逆差矩与患者的生存显著相关($P<0.05$)。具体数据见表3。

对两种分割方法结果的平均数据进行单因素分析发现:径向方差、边缘粗糙度、GLCM差熵、GLCM非相似性、GLCM熵与患者的生存显著相关($P<0.05$)。具体数据见表4。

从表2、表3和表4中可以发现,3种情况下径向方差、边缘粗糙度、GLCM熵和GLCM非相似性与患者的生存情况显著相关,故将这4种特征作为多因素分析的输入协变量。

2.3 多因素分析

将患者的生存时间、现阶段存活状态以及备选

表2 区域生长法分割结果的单因素分析
Tab.2 Univariate analysis of region of interest (ROI) based on region growing method

特征	中位生存时间/月	卡方值	P值
灰度均值			
≥0.899	32.0	4.057	0.044
<0.899	25.5		
径向方差			
≥2.204	20.0	4.530	0.033
<2.204	34.0		
边缘粗糙度			
≥12.339	28.0	4.139	0.042
<12.339	30.5		
GLCM 非相似性			
≥1.261	32.0	4.026	0.045
<1.261	22.0		
GLCM 熵			
≥3.748	37.5	4.539	0.033
<3.748	20.0		

表3 基于边缘检测法分割结果的单因素分析
Tab.3 Univariate analysis of ROI based on edge detection method

特征	中位生存时间/月	卡方值	P值
径向方差			
≥2.213	24.0	5.465	0.019
<2.213	33.0		
边缘粗糙度			
≥11.294	28.0	4.318	0.038
<11.294	30.0		
GLCM 角二阶矩			
≥0.046	23.0	4.014	0.045
<0.046	37.5		
GLCM 差熵			
≥1.496	34.0	6.565	0.010
<1.496	22.0		
GLCM 非相似性			
≥1.261	32.0	6.457	0.034
<1.261	22.0		
GLCM 熵			
≥3.780	32.0	4.098	0.043
<3.780	23.5		
GLCM 逆差矩			
≥0.573	24.0	5.338	0.021
<0.573	31.5		

表4 平均数据的单因素分析
Tab.4 Univariate analysis of average data

特征	中位生存时间/月	卡方值	P值
径向方差			
≥2.253	20.0	6.374	0.012
<2.253	34.5		
边缘粗糙度			
≥12.284	28.0	4.139	0.042
<12.284	30.5		
GLCM 差熵			
≥1.491	34.0	9.528	0.002
<1.491	19.0		
GLCM 非相似性			
≥1.256	32.0	5.909	0.015
<1.256	22.0		
GLCM 熵			
≥3.767	32.0	4.449	0.035
<3.767	22.0		

的协变量按照“向前:LR”的方式输入回归模型,经COX模型多因素分析发现:只有“径向方差”这一影像特征与肺腺癌预后具有显著性关系,即“径向方差”是肺腺癌的独立预后因素。具体数据见表5。

2.4 分类器检验

本文涉及的肺腺癌患者的中位生存时间是29个月,故将现有数据分为存活时间>29个月和≤29个月两类。通过留一交叉检验的方法进行分类实验,实验结果如表6所示。

3 讨论

本研究设计实验对61位肺腺癌患者进行预后因素的分析研究。从图2可以看出,随着存活时间的增加,患者的累积存活率急剧下降,这说明患者的预后质量亟待提高。综合观察表2、表3和表4可以发现:径向方差、边缘粗糙度、GLCM非相似性、GLCM熵对患者的累积生存率的影响有统计学意义(P<0.05)。从表5可以发现两种分割方法的数据以及平均后数据的分析结果都表明径向方差是肺腺癌的独立预后因素。分类器的分类结果显示,径向方差特征对患者生存时间的分类准确率较高,利用该独立预后因素可以对患者的生存时间进行较准确的预后。

在单因素分析实验中,GLCM非相似性与患者预后的显著性关系较为明显,但是在COX模型的多因素分析中,唯有径向方差是肺腺癌预后的独立预后因素。

表5 COX回归模型多因素分析

Tab.5 Multivariate analysis based on COX regression model

ROI 提取方法	特征	B	SE	Wald	P	OR	95%CI
区域生长法	径向方差	0.897	0.439	4.173	0.041	2.452	1.037~5.800
边缘检测法	径向方差	1.010	0.454	4.945	0.026	2.745	1.127~6.685
平均数据	径向方差	1.086	0.455	5.699	0.017	2.961	1.214~7.219

B值:多因素回归分析的回归系数;SE:标准误差;Wald:统计量;P:相关系数;OR值:比值比,OR值大于1证明该因素是危险因素否则为保护因素;95%CI:置信区间

表6 分类实验结果(%)

Tab.6 Classification results(%)

ROI 提取方法	敏感度	特异性	准确率
区域生长法	93.30	32.30	62.30
边缘检测法	83.30	29.00	55.70
平均数据	93.00	12.90	52.46

经分析后,导致这种现象发生的原因可能是不同输入协变量相互影响,不足以产生足够强的相关性。在分类器预后能力检测中,本文的实验结果与国外相关研究的77.5%准确率^[5]相比还有一定差距,但受限于数据总量,用于训练分类器的样本量较少,也是导致分类准确率较国外研究偏低的一个重要因素。

相较于某些国外研究,本文提取了更为全面的图像特征,增加了实验结果验证步骤。而在国内期刊范围内,几乎没有文献表明已经开始了以CT图像特征作为肺腺癌预后因素的相关研究。若能够获取更多可用的图像数据和随访数据,必将筛选出更加准确的预后因素,进而可以对肺腺癌进行更加精准的预后。

本文提取肺腺癌患者CT图像的图像特征,使用Kaplan-Meier方法和COX回归方法对特征数据进行生存分析并使用SVM对结果进行验证发现径向方差这一影像学特征是肺腺癌的独立预后因素。本研究为肺腺癌预后风险评估系统的建立提供了一定的支持。

【参考文献】

- [1] ALBERG A J, BROCK M V, FORD J G, et al. Epidemiology of lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines [J]. Chest, 2013, 143(5 Suppl): e1S-e29S.
- [2] CHEN W Q, ZHENG R S, BAADE P D, et al. Cancer statistics in China, 2015[J]. CA Cancer J Clin, 2016, 66(2): 115-132.
- [3] PASTOR M D, NOGAL A, MOLINA-PINELO S, et al. Identification of proteomic signatures associated with lung cancer and COPD[J]. J Proteomics, 2013, 89(16): 227-237.
- [4] GROVE O, BERGLUND A E, SCHABATH M B, et al. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma [J]. PLoS One, 2015, 10(3): e0118261.
- [5] HAWKINS S H, KORECKI J N, BALAGURUNATHAN Y, et al. Predicting outcomes of non-small cell lung cancer using CT image features[J]. IEEE Access, 2014, 2: 1418-1426.
- [6] BALAGURUNATHAN Y, GU Y H, WANG H, et al. Reproducibility and prognosis of quantitative features extracted from CT images[J]. Transl Oncol, 2014, 7(1): 72-87.
- [7] 杨为贵, 马华玲. 肺腺癌的临床特点及预后分析[J]. 中华胸外科电子杂志, 2015, 2(2): 114-118.
YANG W G, MA H L. Clinical characteristics and prognostic analysis of lung adenocarcinoma [J]. Chinese Journal of Thoracic Surgery (Electronic Edition), 2015, 2(2): 114-118.
- [8] 廉政君, 黄建国. 晚期非小细胞肺癌预后相关因素分析[J]. 现代肿瘤医学, 2014, 22(1): 95-97.
LIAN Z J, HUANG J G. Analysis of factors associated with prognosis of advanced non-small cell lung cancer (NSCLC) [J]. Journal of Modern Oncology, 2014, 22(1): 95-97.
- [9] PRIOR F W, CLARK K, COMMEAN P, et al. TCIA: an information resource to enable open science [C]//Engineering in Medicine and Biology Society. IEEE, 2013: 1282-1285.
- [10] KAMIYA A, MURAYAMA S, KAMIYA H, et al. Kurtosis and skewness assessments of solid lung nodule density histograms: differentiating malignant from benign nodules on CT[J]. Jpn J Radiol, 2014, 32(1): 14-21.
- [11] LEE G, LEE H Y, PARK H, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: state of the art[J]. Eur J Radiol, 2017, 86: 297-307.
- [12] COOK G J, YIP C, SIDDIQUE M, et al. Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? [J]. J Nucl Med, 2013, 54(1): 19-26.
- [13] FRIED D V, TUCKER S L, ZHOU S, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer [J]. Int J Radiat Oncol Biol Phys, 2014, 90(4): 834-842.
- [14] GANESHAN B, PANAYIOTOU E, BURNAND K, et al. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival [J]. Eur Radiol, 2012, 22(4): 796-802.
- [15] ELLISON L F. Estimating relative survival for cancer: an analysis of bias introduced by outdated life tables [J]. Health Rep, 2014, 25(2): 13-19.
- [16] KLEINBAUM D G, KLEIN M. Kaplan-Meier survival curves and the Log-Rank test [M]//Statistics for Biology & Health. Springer, 2012: 45-82.
- [17] ATA N, SÖZER M T. Cox Regression models with nonproportional hazards applied to lung cancer survival data [J]. Hacettepe University Bulletin of Natural Sciences & Engineering, 2014, 36(2): 157-167.

(编辑:薛泽玲)