

CT 三维最小类内散度多分类支持向量机在肺结节识别中的应用

范小波¹, 董莹¹, 王义云² (1. 山东省莱芜市人民医院 271199; 2. 山东省莱芜市牛泉中心卫生院 271124)

【摘要】目的 分析 CT 三维最小类内散度多分类支持向量机(MC-SVM)对肺结节的识别能力及优点。**方法** 选择 2012 年 1 月至 2014 年 1 月确诊的肺结节病患者 50 例, 根据基于三维矩阵模式的感兴趣体(VOI)的构成, 分为结节样和非结节样; 采用自动提取算法提取感兴趣区(ROI), 分为结节 ROI 和非结节 ROI; 采用受试者工作特征(ROC)曲线比较大规模训练人工神经网络(MTANN)、基于矩阵模式的模糊最小二乘 SVM(matFLSSVM)、三维矩阵模式 MC-SVM 和三维最小类内散度 MC-SVM 的识别精度, 同时比较各种方法在不同截断点时的真阳性率和假阳性率。**结果** 三维最小类内散度 MC-SVM 的识别精度、真阳性率均高于其他算法, 而假阳性率低于其他算法($P < 0.05$)。**结论** 三维最小类内散度 MC-SVM 对肺结节的识别精度较高, 值得临床推广应用。

【关键词】 最小类内散度; 多分类支持向量机; 三维矩阵; 肺内结节

DOI:10.3969/j.issn.1672-9455.2015.07.026 文献标志码:A 文章编号:1672-9455(2015)07-0940-02

Application of CT images based 3D minimum within-class scatter multi-class support vector machine in the identification of lung nodules FAN Xiao-bo¹, DONG Ying¹, WANG Yi-yun² (1. Laiwu City People's Hospital, Laiwu, Shandong 271199, China; 2. Health Center of Niuquan, Laiwu, Shandong 271124, China)

【Abstract】Objective To analyze the value of CT images based 3D minimum within-class scatter multi-class support vector machine (MC-SVM) in the identification of lung nodules. **Methods** During Jan. 2012 and Jan. 2014, a total of 50 cases with pulmonary sarcoidosis were enrolled. Lung CT based 3D matrix pattern volume of interest (VOI) structure were divided into nodules samples and non-nodule samples. Region of interest (ROI) was extracted by using automatic extraction algorithm, which was further divided into nodules and non-nodular ROI. Then the recognition accuracy, true positive rate and false positive rate of large-scale massive training artificial neural network (MTANN), fuzzy 'one on one' multi-class SVM (matFLSSVM), 3D matrix model MC-SVM and 3D minimum within-class scatter MC-SVM were compared by using receiver operating curve (ROC). **Results** 3D minimum within-class scatter MC-SVM was with the highest recognition accuracy and true positive rate, and the lowest false positive rate, compared with the other algorithms ($P < 0.05$). **Conclusion** 3D minimum within-class scatter MC-SVM might be accurate for the recognition if long nodules.

【Key words】 minimum within-class scatter; multi-class support vector machine; 3D matrix; lung nodules

肺结节病属于肉芽肿性疾病, 极易侵犯肺、双侧肺门淋巴结及皮肤等多种组织、器官^[1]。计算机辅助诊断(CAD)检查可有效地辅助发现半径过小或与周围组织灰度对比度较低的肺结节^[2]。支持向量机(SVM)是机器学习计算中的最新发现, 主要应用于行为识别及目标检测领域, 可有效提高识别精度^[3]。本研究以肺结节病患者为研究对象, 分析了计算机断层扫描(CT)三维最小类内散度 SVM 对肺结节的识别能力及优点。现将研究结果报道如下。

1 资料与方法

1.1 一般资料 2012 年 1 月至 2014 年 1 月于山东省莱芜市人民医院确诊的肺结节病患者 50 例。根据基于三维矩阵模式的感兴趣体(VOI)的构成, 将 50 例患者分为结节样 31 例、非结节样 19 例。结节样患者包括孤立型 8 例, 与血管粘连型 10 例, 与胸膜粘连型 6 例, 与肺门粘连型 4 例, 磨玻璃影结节 3 例。非结节样患者包括大型血管 10 例, 中型血管 6 例, 小型血管 3 例。所有患者均在 CT 检查后, 通过支气管镜、肺活检或

开胸活检获得病理学明确诊断。

1.2 方法

1.2.1 候选结节提取算法 采用圆点滤波器逐层对 CT 序列进行滤波增强, 提取孤立性结节和粘连型结节; 采用滚球法逐层进行修复, 进一步提取粘连型结节; 在三维空间内进行 26-连接, 构成候选结节的三维 VOI^[1]。

1.2.2 最小类内散度 SVM 定义样本集 $\{y_i^{(k)}, x_i\}_{i=1, k=1}^{n, k=M}$, $y_i^{(k)}$ 表示第 i 个输出中的第 k 个码字, 每个码字定义为 1 和 -1 两种形式, M 个码字构成的输出向量表示 $2M$ 种类别。其中样本子集 D_p 属于第 p ($p=1, 2, \dots, 2^M$) 类, 类内散度 SVM 矩阵定义为 $S = \sum_{p=1}^{2^M} \sum_{x_i \in D_p} (x_i - \bar{x})(x_i - \bar{x})^T$, $\bar{x}$ 为子集 D_p 中所有样本的均值, 最小类内散度矩阵模式的 SVM 优化方程为 $\min 1/2u^T Su + C/2 \sum_{i=1}^n \xi_i^2$ ^[2]。

1.2.3 三维最小类内散度 SVM 1 个三维矩阵可以被展开为 3 个不同方向的二维矩阵 $A_{(1)}, A_{(2)}, A_{(3)}$, 而 $A_{(r)}$ 代表第 r 个三维矩阵的第 r 种展开模式。三维最小类内散度多分类

SVM(MC-SVM)的决策函数如下：

$$g(r)(A_{i_{i_{(r)}}})=[y^{(0)}_{i_{i_{(r)}}}, y^{(1)}_{i_{i_{(r)}}}, \dots, y^{(20)}_{i_{i_{(r)}}}] = \left\{ \begin{array}{l} \left. \begin{array}{l} y_1 \\ y_2 \\ \vdots \\ y_L \end{array} \right\} \rightarrow Y_{C^+} \Rightarrow A_{i_{i_{(r)}}} \in C_{\text{结节}} \\ \left. \begin{array}{l} y_{L+1} \\ \vdots \\ y_{2^w} \end{array} \right\} \rightarrow Y_{C^-} \Rightarrow A_{i_{i_{(r)}}} \in C_{\text{非结节}} \end{array} \right.$$

$r=1,2,3$

其中每个 $y^{(k)}_{i_{i_{(r)}}} = \text{sign}(u^{(k)}_{k,(r)} A_{i_{i_{(r)}}} v_{k,(r)} + b_{k,(r)})$ ，三维矩阵 A 共

产生 3 个决策函数,即 $g(1)、g(2)、g(3)$,实践中,3 人投票选举作出最终决策: $g(A)=g(1) \cap g(2)+g(2) \cap g(3)+g(1) \cap g(3)^{[3]}$ 。

1.3 统计学处理 采用 SPSS18.0 软件进行数据处理和统计学分析。计数资料以百分率表示,两组间比较采用 χ^2 检验。采用受试者操作特征(ROC)曲线比较 4 种 CAD 算法的识别精度。 $P<0.05$ 表示比较差异具有统计学意义。

2 结 果

2.1 4 种 CAD 算法结果比较 4 种 CAD 算法结果比较见表 1。三维最小类内散度 MC-SVM 的真阳性率高于其他算法,而假阳性率低于其他算法($P<0.05$)。

表 1 4 种 CAD 算法结果比较[% (n/n)]

算法	ROI 分类	截断点				
		20%	40%	60%	80%	100%
大规模训练人工神经网络(MTANN)	结节 ROI	100.00(714/714)	95.38(681/714)	92.02(657/714)	91.32(652/714)	89.36(638/714)
	非结节 ROI	14.52(7 011/48 286)	10.28(4 964/48 286)	8.61(4 157/48 286)	7.48(3 612/48 286)	2.97(1 434/48 286)
基于矩阵模式的模糊最小二乘 SVM(matFLSSVM)	结节 ROI	100.00(714/714)	97.17(694/714)	95.10(679/714)	92.86(663/714)	91.18(651/714)
	非结节 ROI	11.96(5 775/48 286)	9.92(4 790/48 286)	7.24(3 496/48 286)	5.97(2 883/48 286)	2.51(1 212/48 286)
三维矩阵模式 MC-SVM	结节 ROI	100.00(714/714)	100.00(714/714)	100.00(714/714)	97.48(696/714)	95.12(679/714)
	非结节 ROI	9.95(4 804/48 286)	7.09(3 423/48 286)	4.79(2 313/48 286)	2.35(1 135/48 286)	1.26(608/48 286)
三维最小类内散度 MC-SVM	结节 ROI	100.00(714/714)	100.00(714/714)	100.00(714/714)	98.46(703/714)	97.48(696/714)
	非结节 ROI	8.94(4 316/48 286)	6.45(3 114/48 286)	4.03(1 946/48 286)	2.96(1 429/48 286)	1.02(493/48 286)

注:ROI 为感兴趣区。

2.2 4 种 CAD 算法的识别精度比较 采用 ROC 曲线比较 4 种 CAD 算法的识别精度结果,MTANN、matFLSSVM、三维矩阵模式 MC-SVM、三维最小类内散度 MC-SVM 曲线下面积分别为 0.975、0.982、0.991、0.996。三维最小类内散度 MC-SVM 识别精度高于其他算法($P<0.05$)。

3 讨 论

肺结节主要是指肺内部类球形且半径 3~30 mm 的病灶^[4]。由于肺结节的半径较小,结节模糊不清,灰度与周围组织的对比度较低,难于被发现,极易被忽略^[5]。肺结节是引发肺部恶性肿瘤的重要因素,也是诊断肺部肿瘤的重要依据。尽早检出肺结节可以有效缓解患者病情的进一步恶化,提高治疗效果^[6]。目前,临床检测肺结节的方法包括 MTANN、Mat-FLSSVM、三维矩阵模式 MC-SVM 及三维最小类内散度 MC-SVM。由于神经网络自身的缺陷,存在对初始值较为敏感、局部最小化、不收敛等问题,直接影响了神经网络方法在 CAD 诊断结节中的应用效果^[7]。三维最小类内散度 MC-SVM 是机器学习方法中最为先进的一种。该方法基于统计学习和结构风险最小化原则,不仅兼顾了学习算法的经验风险与推广能力,有效结合了模型的复杂性与学习能力,还可以转化为 1 个二次型寻优问题,解决了神经网络中局部极值问题,避免了神经网络拓扑结构所需要的经验试凑的方法^[8]。MC-SVM 提出了一种基于相邻断层限制的区域生长法,用于提取候选结节并构成三维 VOI;通过提取所获得的孤立型与浅度粘连型结节,不断

修复相邻层中因正常组织粘连程度较大而被漏提取的候选结节,并将候选结节所在的连续断层内 ROI 构成三维矩阵模式 VOI,为三维矩阵模式的识别奠定了基础。三维最小类内散度 MC-SVM 有效提高了检测肺结节的精确度,为临床治疗提供了可靠的依据^[9]。

本研究以 50 例肺结节病患者为研究对象,比较了 4 种 CAD 算法对肺结节的检测效果及检测价值,并分析了 4 种算法的临床检测特征。结果显示,三维最小类内散度 MC-SVM 的真阳性率高于其他算法,而假阳性率低于其他算法($P<0.05$);MTANN、matFLSSVM、三维矩阵模式 MC-SVM、三维最小类内散度 MC-SVM ROC 曲线下面积分别为 0.975、0.982、0.991、0.996,三维最小类内散度 MC-SVM 的识别精度高于其他算法($P<0.05$)。这表明三维最小类内散度 MC-SVM 对肺结节的检测效果优于其他方法,具有更高的临床检测价值,和相关文献报道的数据相一致^[10]。因此,对于肺结节的诊断,可以采用三维最小类内散度 MC-SVM,以提高疾病的检测效率。

综上所述,三维最小类内散度 MC-SVM 检测肺结节效果明显,其识别精度高于其他检测方法,安全性较高,值得推广。

参考文献

[1] 王青竹,康文伟,王新竹,等.基于三维最小类内散度 SVM 的肺 CT 中的结节识别[J].模式识(下转第 944 页)

3 讨 论

随着非溶血性输血反应日益受到重视,临床广泛应用滤除白细胞的悬浮红细胞。在滤除白细胞的过程中,滤器对脂肪血中红细胞的影响,以及红细胞在储存期的质量,均有待考证。因此,本研究初步探讨了由脂肪血制备而成的滤除白细胞的悬浮红细胞的质量变化,从而为制订合理、有效的降低脂肪血报废率的方法提供实验依据。

目前,血站成分制备人员主要根据肉眼观察血浆外观浑浊程度来判断是否为脂血。已知血浆中 TG 浓度越高,血浆外观越浑浊。因此本研究以 TG 作为筛选脂血的主要指标。为了探讨过滤过程是否对红细胞膜产生影响,首先检测了红细胞渗透脆性。红细胞渗透脆性反映了红细胞在低渗条件下的抵抗力,常作为红细胞膜损伤的指标之一。在脂血过滤后,中、重度组脂血红细胞渗透脆性明显增加,而轻度组、正常组红细胞渗透脆性未见明显变化,提示中、重度组脂血在通过微孔滤膜时,红细胞膜易受到损伤。这与文献报道的高浓度 TG 对红细胞刚性影响较大的结论相一致^[6-7]。另外,王现伟等^[8]在高脂血症大鼠动物模型的研究中也发现了大鼠红细胞在通过微孔膜时,红细胞膜受到严重损伤。深入研究其机制发现,高浓度 TG 能够破坏红细胞骨架完整性,降低红细胞骨架蛋白 F-actin,从而破坏红细胞结构基础,使红细胞变形能力降低,导致红细胞在通过微孔膜时受到挤压损伤。因此,中、重度脂肪血在制备滤除白细胞的悬浮红细胞过程中,红细胞质量受到影响。通过研究红细胞储存期质量变化发现,中、重度组红细胞 K⁺ 浓度虽然在保存第 0 天与轻度组、正常组差异无统计学意义($P>0.05$),但是从保存第 7 天至保存第 35 天,均明显高于轻度组和正常组($P<0.05$),这可能是由于中、重度组红细胞膜受到损伤后,细胞内 K⁺ 外溢程度高于轻度组、正常组所致。检测其游离血红蛋白浓度发现,在保存第 0 天中、重度组游离血红蛋白浓度已经明显高于轻度组、正常组($P<0.05$),且随着保存时间的延长,升高趋势越明显,尤其是在保存末期(第 28~35 天)出现了急剧升高,提示中、重度组红细胞溶血较轻度组、正常组严重。此外,肉眼观察各组红细胞上清液,在保存第 0 天,轻度组、正常组上清液无色透明,而中、重度组上清液呈微红色;至保存第 35 天,轻度组、正常组上清液仅呈微红色,而中、重度组上清液呈明显红色且微浑。这进一步揭示在保存期间中、重度组红细胞寿命较轻度组、正常组大为缩短。在整

个储存期间 Na⁺ 浓度、pH 值各组均呈现下降的趋势,但各组间无明显差异,与吕秋霜等^[9]在对全血、悬浮红细胞保存过程质量研究中所观察到的结果一致。

综上所述,尽管目前国家标准对滤除白细胞的悬浮红细胞的各项生化指标尚无明确规定^[10],但是这种生理功能受到影响的红细胞输注到患者体内后,势必会对患者造成一定的影响。因此,对于轻度脂肪血可采用弃其血浆,利用红细胞,从而达到降低报废率的目的;而对于中、重度脂肪血应该选择全血报废。

参考文献

[1] 金志坚,陈铭娟,姜淑美,等.无偿献血者重度脂肪血调查分析[J].中国输血杂志,2005,17(5):358-358.

[2] 艾新芳,方大治.脂血报废的原因分析及对策[J].中国卫生质量管理,2006,12(5):64-65.

[3] 赵林园,赵树华,朱琳.一种提高脂肪血临床利用率的脂肪血分离方法[J].中国输血杂志,2006,18(5):399.

[4] 谢玉萍,曾常爱.吉安市脂肪血报废的原因分析及利用[J].井冈山医学学报,2007,13(6):54.

[5] 叶应妩,王毓三,申子瑜.全国临床检验操作规程[M].3版.南京:东南大学出版社,2006.

[6] 王梅,耿庆信,刘国梁.血清甘油三酯及胆固醇对血液黏度和红细胞流变性的影响[J].山东医药,2005,45(16):43-44.

[7] 杨菲,贾芳,许进.2 型糖尿病大鼠血清甘油三酯,红细胞聚集性动态变化及其相关性[J].山东医药,2007,47(5):24-25.

[8] 王现伟,胡林,许锋林,等.激光生物学技术研究高脂血症大鼠红细胞变形能力改变及其机制[J].激光生物学报,2007,16(2):148-151.

[9] 吕秋霜,任芙蓉,刘长利.4 种红细胞制剂在保存过程中的质量观察[J].中国输血杂志,2008,20(6):506-507.

[10] 周静宇.全血及成分血质量要求(GB18469-2012)实施探讨[J].临床血液学杂志:输血与检验,2013,26(8):572-573.

(收稿日期:2014-09-15 修回日期:2014-12-19)

(上接第 941 页)

别与人工智能,2011,24(5):700-706.

[2] 王仲楚,方克灶,苏缪广,等.16 层螺旋 CT 诊断肺内结节临床分析[J].医学影像学杂志,2013,23(4):627-628.

[3] 单立群,胡坚,李明东,等.胸腔镜在肺部微小结节诊治中的应用[J].中国肺癌杂志,2013,16(7):369-372.

[4] Sengur A. Multiclass least-squares support vector machines for analog modulation classification[J]. Exp System Applica,2009,36(3):6681-6685.

[5] 史华莉,孙琼芳,涂蓉,等. CT 三维重建技术对肺内孤立性结节的诊断价值[J].中华临床医师杂志:电子版,2014,8(5):88-93.

[6] 张梦龙,孙峰,王雅莹,等.数字合成体层成像和 X 线数字摄影在肺内结节探测中敏感度的比较[J].医学影像学杂

志,2013,23(4):597-600.

[7] 向林峰,文石兵,夏发明.原发周围型肺内孤立小结节的微创诊治[J].临床肺科杂志,2011,16(6):961.

[8] Gangeh MJ, Sørensen L, Shaker SB, et al. A texton-based approach for the classification of lung parenchyma in CT images[J]. Med Image Comput Comput Assist Interven, 2010,13(3):595-602.

[9] 张文玉,黄勇,申洪明,等.肺癌患者肺内小结节良恶性的 CT 诊断价值[J].临床放射学杂志,2011,30(11):1604-1608.

[10] 颜志红,于武江,万云飞. CT 扫描对诊断肺癌肺内转移的临床意义[J].实用医学影像杂志,2011,12(2):79-81.

(收稿日期:2014-10-15 修回日期:2014-12-20)