

Evaluation of high-grade prostate cancer with machine learning model of prostate imaging reporting and data system version 2

ZENG Xiaohui, PENG Tao*, GAO Yueqin, NIU Xiangke, CHEN Xuehui,
ZHANG Shihui, CHEN Zhifan

(Department of Radiology, Affiliated Hospital of Chengdu University,
Chengdu 610081, China)

[Abstract] **Objective** To establish the support vector machine (SVM), decision tree (DT) and Logistic regression models with prostate imaging reporting and data system version 2 (PI-RADS v2), and to evaluate their diagnostic efficiency in high-grade prostate cancer. **Methods** Clinical and MRI data of histopathologically proved prostate cancer in 194 patients were retrospectively analyzed, including 63 patients of high-grade cancer group and 131 patients of non high-grade cancer group. The evaluation factors, including PI-RADS v2 score, age, free prostate specific antigen, prostate specific antigen ratio and adjusted-prostate-specific antigen density were analyzed with SVM, DT and Logistic models, and the values of PI-RADS v2 and three models in diagnosis of high-grade prostate cancer were evaluated with ROC. **Results** The sensitivity of PI-RADS v2, SVM, DT and Logistic models in diagnosis of high-grade prostate cancer was 72.73%, 69.09%, 87.27% and 70.91%, while the specificity was 87.29%, 93.22%, 93.22% and 95.76%, respectively. AUC of DT model in diagnosis of high-grade prostate cancer was the largest ($AUC = 0.90$, $P < 0.01$). There were statistically significant differences of AUC of DT and the PI-RADS v2 score, SVM, and Logistic models (all $P < 0.05$). **Conclusion** PI-RADS v2, SVM, DT and Logistic models has good value in diagnosis of high-grade prostate cancer.

[Key words] Prostatic neoplasms; Support vector machine; Decision tree; Logistic models

DOI:10.13929/j.1003-3289.201808198

基于前列腺影像报告和数据系统第 2 版的机器学习模型诊断高级别前列腺癌

曾小辉, 彭涛*, 高月琴, 牛翔科, 陈雪卉, 张仕慧, 陈志凡

(成都大学附属医院放射科, 四川 成都 610081)

[摘要] **目的** 建立基于前列腺影像报告和数据系统第 2 版(PI-RADS v2)的支持向量机(SVM)、决策树(DT)和 Logistic 回归 3 种机器学习模型,评价上述模型对高级别前列腺癌的诊断价值。**方法** 回顾性分析于我院接受前列腺多参数 MR 扫描并取得病理结果的 194 例患者的资料,其中高级别癌 63 例,非高级别癌 131 例。将评价因素(PI-RADS v2 评分、年龄、游离前列腺特异抗原、前列腺特异性抗原比值、前列腺特异抗原密度)录入 SVM、DT 和 Logistic 回归 3 种机器学习模型进行诊断,通过 ROC 曲线评价 PI-RADS v2 评分和 3 种机器学习模型诊断高级别前列腺癌的价值。**结果** PI-RADS v2、SVM、DT 和 Logistic 回归模型诊断高级别前列腺癌的敏感度分别为 72.73%、69.09%、87.27%和 70.91%;特异度分别为 87.29%、93.22%、93.22%和 95.76%。DT 模型诊断高级别前列腺癌 ROC 的 AUC ($AUC = 0.90$, $P <$

[基金项目] 四川省卫计委科研项目(18PJ150)、成都市卫计委科研项目(2015080)。

[第一作者] 曾小辉(1976—),男,四川中江人,本科,主治医师。研究方向:腹部及胸部影像学。E-mail: zxx_njrmc@163.com

[通信作者] 彭涛,成都大学附属医院放射科,610081。E-mail: 474488159@qq.com

[收稿日期] 2018-08-30 **[修回日期]** 2018-11-05

0.01)最大,且与 PI-RADS v2 评分、SVM、Logistic 回归比较差异均有统计学意义(P 均 ≤ 0.05)。结论 PI-RADS v2 评分、SVM、DT 和 Logistic 回归模型诊断高级别前列腺癌的价值均较好。

[关键词] 前列腺肿瘤;支持向量机;决策树;Logistic 模型

[中图分类号] R737.25; R445.2 [文献标识码] A [文章编号] 1003-3289(2018)12-1852-05

前列腺癌居恶性肿瘤导致男性死亡的第 2 位^[1]。随着人民生活质量提高和人口老龄化,我国前列腺癌发病率呈上升趋势^[2]。前列腺多参数 MRI 是目前诊断前列腺癌的主要无创检查方法。为提高前列腺癌诊断率和报告的规范化,2012 年欧洲泌尿生殖放射协会提出前列腺影像报告和数据系统第 1 版(prostate imaging reporting and data system version 1, PI-RADS v1),但存在诸多不足,2014 年底由美国放射学会、欧洲泌尿生殖放射学会及 AdMeTech 基金会共同对 PI-RADS v1 进行修订,在线颁布了前列腺影像报告和数据系统第 2 版(prostate imaging reporting and data system version 2, PI-RADS v2)^[3-4]。随着对前列腺癌的深入研究,有学者^[5-6]发现高级别前列腺癌(Gleason 评分 ≥ 7 分)和非高级别癌患者预后明显不同,但通过多种模型结合影像学表现及临床相关指标诊断高级别前列腺癌的报道较少见。本研究探讨支持向量机(support vector machine, SVM)、决策树(decision tree, DT)和 Logistic 回归 3 种机器学习模型在诊断高级别前列腺癌中的价值。

1 资料与方法

1.1 一般资料 回顾性分析 2013 年 1 月—2018 年 2 月于我院接受前列腺多参数 MR 扫描并取得病理结果的 194 例患者的资料,年龄 50~90 岁,平均(74.7 $\pm$ 8.9)岁。MRI 显示中位前列腺体积为 55.16(35.47, 77.77) cm^3 ;中位总前列腺特异抗原(total prostate specific antigen, tPSA)为 16.30(8.32, 46.52) ng/ml ,中位游离前列腺特异抗原(free prostate specific antigen, fPSA)为 2.08(1.23, 4.65) ng/ml ,前列腺特异性抗原密度(adjusted-prostate-specific antigen density, adjusted-PSAD)为 0.28(0.14, 1.16) $\text{ng}/(\text{ml} \cdot \text{cm}^3)$ 。根据 Gleason 评分^[7]将患者分为 2 组,高级别癌组 63 例(Gleason 评分 ≥ 7 分),Gleason 评分分别为:3+4=7 分 10 例、4+3=7 分 5 例、4+4=8 分 13 例、3+5=8 分 1 例、5+4=9 分 8 例、4+5=9 分 14 例、5+5=10 分 12 例;非高级别癌组 131 例,Gleason 评分 < 7 分。纳入标准:①临床资料及相关检查完整;②MR 检查前 3 个月内未接受前列腺穿刺、手术及前列腺癌内分泌治疗;③MR 检查后

1 个月内经直肠超声引导下穿刺或前列腺全切术,并取得病理结果。排除标准:①图像质量不佳;②MR 序列不完整;③既往有前列腺相关手术或治疗病史。

1.2 仪器与方法 采用 Siemens Avanto 1.5T 超大型 MR 扫描仪,相控阵体线圈,扫描前嘱患者排空膀胱。仰卧位扫描,先行常规 T2WI 及 DWI,后行动态增强扫描。T2WI 参数:FOV 200 mm $\times$ 200 mm,扫描层数 20 层,层厚 3 mm,层间距 0.6 mm,采用自旋回波-回波平面成像序列,TR 4 300 ms,TE 101 ms,矩阵 256 $\times$ 205;DWI:b 值取 800 s/mm^2 ,FOV 260 mm $\times$ 220 mm,扫描层数 20 层,层厚 3.6 mm,层间距 0,TR 3 700 ms,TE 80 ms,矩阵 160 $\times$ 100,NEX 6 次。动态增强扫描时,采用高压注射器经肘前静脉注射 Gd-DTPA 0.2 mmol/kg 体质量,流率 2.5 ml/s;3D-Flash-T1WI,FOV 260 mm $\times$ 260 mm,扫描层数 20 层,层厚 3.6 mm,体块(slab)为 1,TR 5 ms,TE 1.69 ms,矩阵 192 $\times$ 138,扫描次数 36 次。

1.3 相关指标收集及处理 通过临床信息系统(clinical information system, CIS)查阅获得临床指标,包括年龄、fPSA、前列腺抗原比值(fPSA/tPSA)和 adjusted-PSAD。由 2 名具有 5 年工作经验的、泌尿系统影像学诊断医师利用盲法对患者进行 PI-RADS v2 评分^[7],如有分歧经协商达成一致。在 T2WI 上结合轴位、冠状位和矢状位图像测量前列腺左右径、前后径及上下径,根据公式计算前列腺体积(V)及 adjusted-PSAD: $V = \text{左右径}(\text{cm}) \times \text{前后径}(\text{cm}) \times \text{上下径}(\text{cm}) \times 0.52$, $\text{adjusted-PSAD} = \text{tPSA}/V$ 。

1.4 模型建立 采用 IBM SPSS Modeler 软件中的 SVM、DT 和 Logistic 回归分别对高级别前列腺癌建立预测模型,数据流示意图见图 1。对各评价因素和相关诊断数据建立数据库,录入后建立 SVM、DT 和 Logistic 回归 3 种模型;对 5 个评价因素(年龄、PI-RADS v2 评分、fPSA、fPSA/tPSA 及 adjusted-PSAD)建立 SVM 流、DT 流和 Logistic 回归数据流,并分析 3 种学习模型诊断高级别前列腺癌的准确率。采用 IBM SPSS Modeler 软件随机抽样选出 173 例(高级别前列腺癌组 55 例、非高级别前列腺癌组 118 例)作为训练集,用于筛选变量和建立预测模型;21 例

(高级别前列腺癌组 8 例、非高级别前列腺癌组 13 例)作为测试集,用于评价模型预测效果。

1.5 统计学分析 采用 SPSS 20.0 统计分析软件。对数据进行正态分析 (Kolmogorov-Smirnov 检验) 和方差齐性 (Levene 检验) 检验, 计量资料符合正态分布以 $\bar{x} \pm s$ 表示, 非正态分布以中位数 (上下四分位数) 表示。绘制 ROC 曲线评价 PI-RADS v2 评分和 3 种机器学习模型诊断高级别前列腺癌的效能; 并以 DeLong 方法比较 AUC; 以病理 Gleason 评分结果为金标准, 采用 χ^2 检验比较 PI-RADS v2 评分及 3 种模型诊断高级别前列腺癌的敏感度、特异度及准确率的差异。 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 模型评价及数据 以 PI-RADS v2 评分预测高级别前列腺癌的准确率为 82.66%。训练集 SVM、DT 和 Logistic 回归 3 种机器学习模型经学习后诊断高级别前列腺癌的准确率、敏感度及特异度见表 1, PI-RADS v2 评分及 3 种机器学习模型诊断高级别前列腺癌的敏感度、特异度、准确率差异均无统计学意义 (P 均 > 0.05)。测试集 SVM、DT 和 Logistic 回归 3 种机器学习模型经学习后诊断高级别前列腺癌准确率分别为 76.19%、80.95% 及 76.19%。通过 IBM SPSS Modeler 软件经训练集学习后, 自动获得 3 种学习模型诊断高级别前列腺癌变量的重要性, 并按所占比例进行排序: SVM 模型前 3 位预测变量为 PI-RADS v2、adjusted-PSAD 及年龄, 所占比例分别为 50.00%、15.00% 及 13.00%; DT 模型前 3 位预测变量为 adjusted-PSAD、PI-RADS v2 及 fPSA, 所占比例分别为 49.00%、23.00% 及 14.00%; Logistic 回归模型前 3 位预测变量为 PI-RADS v2、fPSA/tPSA 及 adjusted-PSAD, 所占比例分别为 36.00%、27.00% 及 21.00%。

2.2 ROC 曲线分析 173 例训练集 PI-RADS v2 评分和 SVM、DT 和 Logistic 回归 3 种机器学习模型诊断高级别前列腺癌的 AUC 分别为 0.83、0.81、0.90

表 1 测试集 3 种机器学习模型及 PI-RADS v2 评分诊断高级别前列腺癌的比较分析 (%)

项目	敏感度	特异度	准确率
SVM 模型	69.09	93.22	85.55
Logistic 模型	70.91	95.76	87.86
DT 模型	87.27	93.22	91.33
PI-RADS v2 评分	72.73	87.29	82.66
χ^2 值	6.085	6.495	6.111
P 值	0.108	0.090	0.106

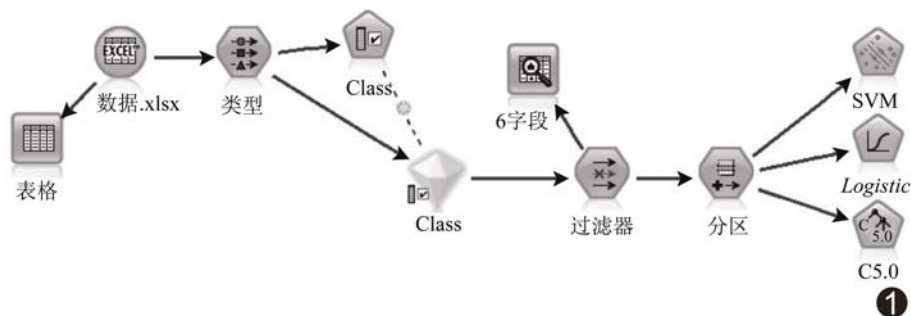


图 1 3 种机器学习模型数据流示意图

及 0.83 (P 均 < 0.01), 其中 DT 模型的 AUC 最大 (图 2), 且与 PI-RADS v2 评分 ($Z = 1.96$, $P = 0.049$)、SVM ($Z = 2.87$, $P = 0.004$) 和 Logistic 回归 ($Z = 2.31$, $P = 0.021$) 差异均有统计学意义, 其模型树状图见图 3; 而 SVM 与 Logistic 回归 ($Z = 1.25$, $P = 0.21$) 及 PI-RADS v2 评分 ($Z = 0.98$, $P = 0.32$)、Logistic 回归与 PI-RADS v2 评分 ($Z = 0.04$, $P = 0.96$) 的 AUC 差异均无统计学意义。

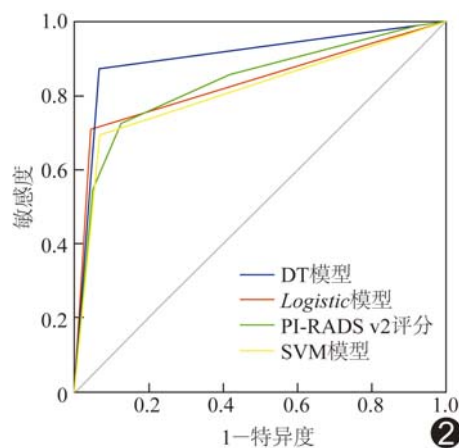


图 2 PI-RADS v2 评分及 3 种机器学习预测高级别前列腺癌的 ROC 曲线

3 讨论

多参数 MRI^[8]是目前诊断前列腺癌的主要影像学

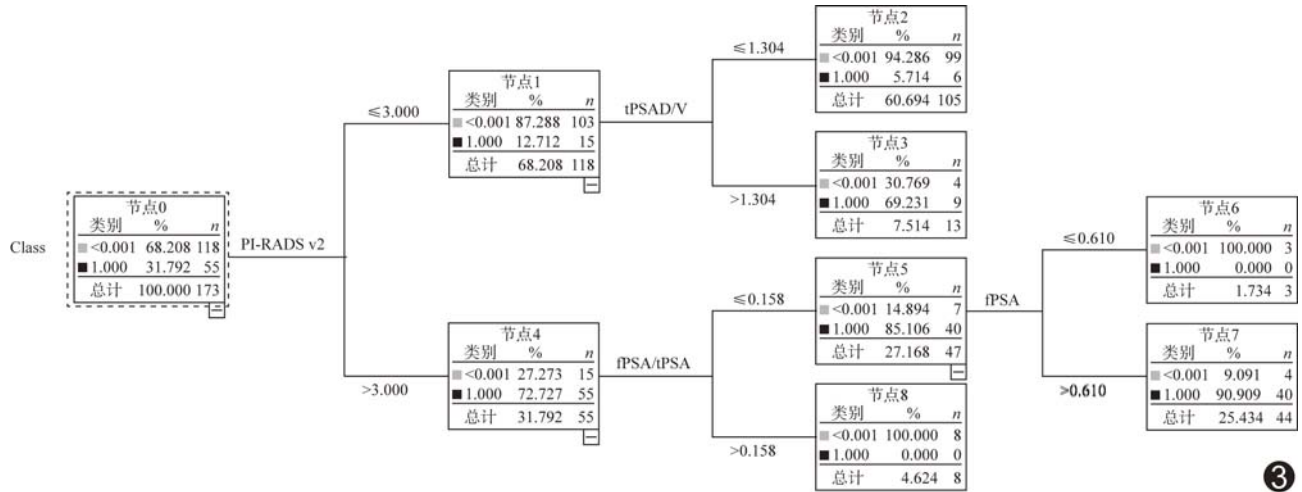


图3 DT模型树状图

检查方法,有助于提高诊断前列腺癌的准确率,但影像学报告不规范等问题限制了其广泛应用。与 PI-RADS v1 相比,PI-RADS v2 评分系统主要针对临床显著性癌或高级别肿瘤^[9],其将临床显著性高级别肿瘤前列腺癌定义为符合以下条件之一者:①前列腺穿刺活检任何一针 Gleason 评分 >6 分;②前列腺穿刺活检任何一针癌组织长度大于 5 mm。Albertsen 等^[10]回顾性分析发现高级别前列腺癌患者 10 年死亡率高于非高级别前列腺癌,并提出应着重诊断高级别前列腺癌,以便临床减少不必要的前列腺穿刺活检及降低对非高级别前列腺癌的过度治疗。De Visschere 等^[11]研究报道,应用单一评分系统对前列腺癌进行预测具有一定风险。本研究联合应用 3 种机器学习模型对前列腺癌影像学表现和临床指标进行分析。

机器学习模型能根据疾病与相关危险因素的内在规律近似地以数学方式表达,已成为预测疾病的主流方法。本研究采用 SVM、DT 和 Logistic 回归建立对高级别前列腺癌的预测模型。SVM 的优点为解决小样本、非线性及高维模式识别问题,追求现有信息情况下的最优解,可提高学习方法的推广能力;DT 具备对未来新数据的分类预测能力,可将大量数据有目的地分类,从中找到潜在的、有价值的信息,且准确性良好,同时还可揭示重要预测因子之间的复杂关系;Logistic 回归的优点是能较好地表现评价因素与因变量之间的依存关系。

本研究中,DT 模型诊断高级别前列腺癌的 AUC 最大,且与 PI-RADS v2 评分、SVM 和 Logistic 回归的 AUC 差异均有统计学意义,而 SVM 与 Logistic 回归及 PI-RADS v2 评分、Logistic 回归与 PI-RADS v2

评分的 AUC 差异均无统计学意义。DT 模型预测高级别前列腺癌诊断准确率、敏感度及特异度均较高,但与其他模型相比差异无统计学意义,可能与阳性样本量较少有关。IBM SPSS Modeler 软件建立的 3 种预测高级别前列腺癌模型前 3 位变量中均有 PI-RADS v2 评分,且单纯采用 PI-RADS v2 评分的诊断价值也较高,与模型预测变量重要性基本相符。van Leeuwen 等^[12]发现运用前列腺癌相关临床指标联合 PI-RADS v1 评分建立 Logistic 模型预测前列腺高级别癌的 AUC 为 0.88。Bermejo 等^[13]认为基于前列腺特异性抗原等单独生物指标不能很好地诊断前列腺癌,故联合年龄、PSA、直肠指检、多参数 MR 等指标建立 DT 和 Logistic 回归模型,发现 2 种模型诊断前列腺癌的准确率均较好(AUC=0.72,0.80)。Nhun 等^[14]采用 SVM 模型预测前列腺癌的敏感度为 95.1%、特异度为 84.6%。上述研究^[12-14]的目的均在于鉴别诊断前列腺癌与良性前列腺病变,而对高级别与非高级别前列腺癌的分类研究较为缺乏,且上述研究均未对各模型进行对比分析。Lee 等^[15]基于经直肠超声,联合临床相关指标,建立了诊断前列腺癌的 Logistic 回归、人工神经网络和 SVM 模型,发现 SVM 模型的 AUC 最大(AUC=0.847),且与其他 2 种模型比较差异有统计学意义。本研究结果与其不符,分析原因:①研究纳入的决策因素不同;②Lee 等^[15]的研究基于超声,而本研究是基于多参数 MR 分析建立的模型;③本研究样本量较小,SVM 比人工神经网络更为适用。

本研究的不足:①为回顾性研究,前列腺大切片病理标本及穿刺点不能完全与影像学直接对照,结果可

能存在偏倚;②样本量小,对于样本选择有一定局限性;③仅为 3 种机器学习模型的训练集和较少测试集分析,有待样本量扩大后进一步验证。

综上所述,基于 PI-RADS v2 评分的 3 种算法模型预测高级别前列腺癌的价值均较好,DT 模型可作为预测前列腺高级别癌的较优模型。

[参考文献]

- [1] Godley KC, Syer TJ, Toms AP, et al. Accuracy of high b-value diffusion-weighted MRI for prostate cancer detection: A meta-analysis. *Acta Radiol*, 2018,59(1):105-113.
- [2] Hassanipour-Azgomi S, Mohammadian-Hafshejani A, Ghoncheh M, et al. Incidence and mortality of prostate cancer and their relationship with the Human Development Index worldwide. *Prostate Int*, 2016,4(3):118-124.
- [3] Junker D, Quentin M, Nagele U, et al. Evaluation of the PI-RADS scoring system for mpMRI of the prostate: A whole-mount step-section analysis. *World J Urol*, 2015,33(7):1023-1030.
- [4] Park SY, Jung DC, Oh YT, et al. Prostate cancer: PI-RADS Version 2 helps preoperatively predict clinically significant cancers. *Radiology*, 2016,280(1):108-116.
- [5] Paterson NR, Lavallée LT, Nguyen LN, et al. Prostate volume estimations using magnetic resonance imaging and transrectal ultrasound compared to radical prostatectomy specimens. *Can Urol Assoc J*, 2016,10(7-8):264-268.
- [6] 郭雪梅,王霄英,吴冰,等.前列腺外周带癌的 ADC 值与其病理分级和临床分期的相关性. *中国医学影像技术*, 2009, 25(4): 674-677.
- [7] Epstein JI, Allsbrook WC, Amin MB, et al. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol*, 2005,29(9):1228-1242.
- [8] 李勇,娄明武.多参数磁共振成像技术早期诊断前列腺癌研究进展. *中国医学影像技术*, 2017,33(2):307-311.
- [9] Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging-Reporting and Data System: 2015, Version 2. *Eur Urol*, 2016,69(1):16-40.
- [10] Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA*, 2005,293(17):2095-2101.
- [11] De Visschere PJ, Briganti A, Fütterer JJ, et al. Role of multiparametric magnetic resonance imaging in early detection of prostate cancer. *Insights Imaging*, 2016,7(2):205-214.
- [12] van Leeuwen PJ, Hayen A, Thompson JE, et al. A multiparametric magnetic resonance imaging-based risk model to determine the risk of significant prostate cancer prior to biopsy. *BJU Int*, 2017,120(6):774-781.
- [13] Bermejo P, Vivo A, Tárraga PJ, et al. Development of interpretable predictive models for BPH and prostate cancer. *Clin Med Insights Oncol*, 2015,9:15-24.
- [14] Nhung NT, Khuong VT, Huy VQ, et al. Classifying prostate cancer patients based on total prostate-specific antigen and free prostate-specific antigen features by support vector machine. *J Cancer Res Ther*, 2016,12(2):818-825.
- [15] Lee HJ, Hwang SI, Han SM, et al. Image-based clinical decision support for transrectal ultrasound in the diagnosis of prostate cancer: Comparison of multiple Logistic regression, artificial neural network, and support vector machine. *Eur Radiol*, 2010,20(6):1476-1484.

使用阿拉伯数字和汉字数字的一般原则

根据 GB/T 15835《出版物上数字用法的规定》

(1)在统计图表、数学运算、公式推导中所有数字包括正负整数、小数、分数、百分数和比例等,都必须使用阿拉伯数字。

(2)在汉字中已经定型的词、词组、成语、缩略语等都必须使用汉语数字,例如:一次方程、三维超声、二尖瓣、法洛四联症、星期一、五六天、八九个月、四十七八岁等。

(3)除了上述情况以外,凡是使用阿拉伯数字而且又很得体的地方,都应该使用阿拉伯数字。遇到特殊情况时,可以灵活掌握,但应该注意使全篇同一。

(4)如果数字的量级小于 1 时,小数点前面的零(0)不能省去,如 0.32 不能写成.32。