

## 基于Web的剂量体积直方图数据自动提取工具开发及验证

吴昊<sup>1</sup>, 刘杰<sup>2</sup>, 岳海振<sup>1</sup>, 蒋璠<sup>1</sup>, 张艺宝<sup>1</sup>

1. 北京大学肿瘤医院暨北京市肿瘤防治研究所放疗科, 恶性肿瘤发病机制及转化研究教育部重点实验室, 北京 100142; 2. 北京东方瑞云科技有限公司, 北京 100020

**【摘要】目的:**开发并测试一套基于Web的剂量体积直方图(DVH)数据自动提取工具。**方法:**(1)采用Django应用框架和Python编程语言,设计一套基于Web的DVH数据自动提取工具。(2)利用自动工具分析从Eclipse计划系统中导出的30例相同类型计划的DVH表单数据,并采用人工方法读取DVH图中的相应参数作为对比,分析其耗时、准确性等方面的表现及误差产生原因。**结果:**自动提取DVH数据的效率远高于人工分析,且正确率更有保障。对于计划靶区体积的均匀性指数、股骨头和膀胱D<sub>50%</sub>和平均剂量等参数,自动与人工提取的差异极小(误差 $\leq 0.01\%$ ,  $P > 0.05$ )。但对于适形指数(CI)值的计算,由于计划系统空间采样算法的原因使得基于等剂量线结构转换测量以及DVH表单数据分析之间的结果存在较大差异[CI<sub>PGTV</sub>平均相差(2.60 $\pm$ 1.04)%, CI<sub>PTV</sub>平均相差(0.66 $\pm$ 0.29)%,  $P < 0.001$ ],但Web工具采用的DVH分析结果更加接近Eclipse自动生成的CI值,且有效避免了后者一次只能计算一个CI值的缺陷。**结论:**本工作开发的基于Web的工具可以对海量DVH数据进行高效、准确的自动统计,且具有跨平台应用等优势。

**【关键词】**Web; 剂量体积直方图; 自动分析; 大数据; 放疗计划; 剂量学

**【中图分类号】**R730.55; R319

**【文献标识码】**A

**【文章编号】**1005-202X(2016)05-0433-04

## Development and validation of Web-based automated dose-volume histogram data analyzer

WU Hao<sup>1</sup>, LIU Jie<sup>2</sup>, YUE Hai-zhen<sup>1</sup>, JIANG Fan<sup>1</sup>, ZHANG Yi-bao<sup>1</sup>

1. Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiotherapy, Peking University Cancer Hospital, Beijing Institute for Cancer Research, Beijing 100142, China; 2. Beijing Eastraycloud Technology Co. Ltd. Beijing 100020, China

**Abstract: Objective** To develop and validate a Web-based automated dose-volume histogram (DVH) data analyzer. **Methods** Based on the Django framework and Python programming language, a web-based automated DVH data analyzer was designed. The DVH in tabular format of 30 similar plans were exported from Eclipse planning system by using automated analyzer; the corresponding parameters of DVH were extracted manually. And the exported data were compared with the extracted data to analyze the efficiency, accuracy and the reasons of errors. **Results** Compared with manual method, the automated DVH data analyzer showed superior efficiency and accuracy. For the parameters such as HI<sub>PTV</sub>, D<sub>50%</sub> and mean dose for the femoral head and urinary bladder, the disparities between manual and automatic extraction were negligible (Error $\leq 0.01\%$ ,  $P > 0.05$ ). However, significant differences in the calculation of conformity index (CI) were observed between the isodose conversion method and the DVH extraction, which was ascribed to the spatial sampling algorithms of planning system. The mean difference in CI<sub>PGTV</sub> and CI<sub>PTV</sub> were respectively (2.60 $\pm$ 1.04)% and (0.66 $\pm$ 0.29)%,  $P < 0.001$ . The results of DVH extraction adopted by the Web-based analyzer were more close to the CI values auto-generated by Eclipse. And the Web-based analyzer effectively conquered the limitation of Eclipse which only calculated a CI value at a time. **Conclusion** The developed Web-based DVH analyzer can auto-analyze large amount of DVH data efficiently and accurately, with the advantage of cross-platform application.

**Key words:** Web dose-volume histogram; auto-analysis; big data; treatment planning; dosimetry

**【收稿日期】**2016-01-06

**【基金项目】**国家自然科学基金(11505012, 81402535);北京市医院管理局“青苗”计划专项经费资助(QML20151004);质检公益性行业科研专项(201510001-02)

**【作者简介】**吴昊(1976-), 硕士, 高级工程师, 研究方向: 医学物理, Tel: 010-88196033, E-mail: wuhao\_bicr@sina.com。

**【通信作者】**张艺宝(1985-), 博士, 工程师, 研究方向: 医学物理, Tel: 010-88196033, E-mail: ybzhang66@163.com。

## 前言

剂量体积直方图(Dose-Volume Histogram, DVH)不仅是制定和优化患者放疗方案的重要依据<sup>[1-3]</sup>,同时也是进行预后判断<sup>[4-5]</sup>、质量控制<sup>[6-7]</sup>、剂量学分析<sup>[8-10]</sup>、多中心比对<sup>[11-12]</sup>的标准方法之一。商用放疗计划系统提供了多种工具用于分析DVH数据,主要分为曲线读点和参数计算两类<sup>[13]</sup>。其中曲线读点方式相对耗时且存在测量者视觉误差,而参数计算只能实现厂家预先设定的基本功能。虽然两类工具的结合可以满足当前大部分临床个案分析需求,但随着大数据时代的来临以及海量病例分析需求的增加,人工读取和数据誊写的低效率和高错误率等问题日益显现。本工作以直肠癌术前同步推量计划为例,开发了一套基于Web的DVH参数自动提取工具,并以手工提取的30套DVH数据作为参照进行了对比验证。

## 1 材料与方法

### 1.1 提取DVH参数

将30例直肠癌术前同步推量计划的DVH数据以表单形式从Eclipse计划系统中导出,并参照国际辐射单位和计量委员会ICRU 83号报告<sup>[14]</sup>以及本院临床习惯,尝试自动提取如下参数:“Body”的 $V_{5060\text{ cGy}}$ 和 $V_{4180\text{ cGy}}$ [该结构被相应等剂量线包含的绝对体积,5060 cGy和4180 cGy分别对应100% PGTV (Planning Gross Target Volume)和PTV (Planning Target Volume)的处方剂量];PGTV和PTV的绝对体积及其 $D_{2\%}$ ,  $D_{50\%}$ ,  $D_{95\%}$ 和 $D_{98\%}$ (2%, 50%, 95%和98%该结构体积所受最低剂量,单位cGy);PGTV中超过107%处方剂量的“热点”体积 $V_{107\%\text{ PGTV}}$ ( $V_{107\%}$ 即 $V_{5414\text{ cGy}}$ );股骨头和膀胱的 $D_{50\%}$ ( $D_{50\%\text{ FH}}$ 、 $D_{50\%\text{ UB}}$ )和平均剂量( $D_{\text{mean\_FH}}$ 、 $D_{\text{mean\_UB}}$ ,单位cGy),并利用以上数据分别计算PGTV和PTV的剂量均匀性指数(Homogeneity Index, HI)<sup>[15]</sup>和适形指数(Conformity Index, CI)<sup>[16]</sup>:

$$HI = (D_{2\%} - D_{98\%}) / D_{50\%}, CI = V_{100\%} / V_{\text{PGTV}} \text{ or } V_{\text{PTV}}$$

### 1.2 程序架构和算法

程序基于Web架构设计,采用Django应用框架和Python编程语言,通过浏览器实现DVH表单数据的批量文件上传,利用服务器后台程序对文件进行智能解析、信息抽取和统计分析,并按预先设定的格式模板以Excel文档形式自动生成统计结果并下载至本地保存。其算法是:程序从DVH表头文件中搜索患者ID、疗程/计划ID并分别作为输出文件的第1列和第2列。由于Eclipse系统仅允许定义一个处方剂量,而对于同步推量计划不同物理师可能输入不同的优化处方。为避免

造成混乱,算法直接对结构和绝对剂量进行定义。在各器官结构数据的开始部分,通过搜索结构名、体积( $\text{cm}^3$ )、平均剂量(cGy)等关键词直接获得相应数据并输出,若平均剂量的单位为%,则在表头中搜索处方剂量用于绝对剂量转换。对于同一结构在不同计划中并存的多种命名规则,算法采用穷举所有可能名称的方式建立相应数据库并从中进行判断匹配。在组成表单DVH主数据的3列结构中,除第3列固定为相对体积外,绝对剂量(cGy)和相对剂量(%)的次序会因不同用户的DVH导出设置而变化。鉴于此,算法通过对首行字符进行逻辑判断,实现对绝对剂量和相对剂量两列数据的识别和利用,再根据需要由已知的总体积计算某剂量对应的绝对体积。考虑到表单DVH数据文件的大小及其导出和上传所耗费的时间,临床上可能依据不同的精度需求设置不同的剂量采样间距。因此,对于不能直接找到的绝对剂量点,算法分别对剂量和体积采用临近两点线性插值的方法进行补充。

### 1.3 作为参照的人工数据提取

利用Eclipse系统提供的剂量参数统计表以及DVH读取工具(如十字叉,放大缩小,等剂量线结构转换及体积测量)等,对上述30例计划的相同数据进行人工采集,记录耗费时间并与Web提取方法进行比较。

### 1.4 统计学方法

基于SPSS 21.0软件包,利用Shapiro-Wilk方法对数据做正态性检验,对于符合及不符合正态分布的数据分别进行配对 $t$ 检验和Wilcoxon秩和检验以判断两种方法提取的数据之间是否具有显著差异(双侧,  $P < 0.05$ ),在对自动和人工提取的数据进行逐项匹配后计算差值并判断临床意义。

## 2 结果

系统开发的整体架构和算法逻辑分别如图1和图2所示。

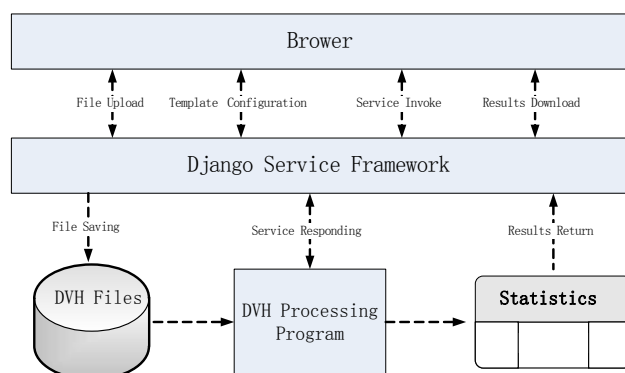


图1 基于Web的DVH数据提取程序架构

Fig.1 Program framework of Web-based DVH data analyzer

DVH: Dose-volume histogram

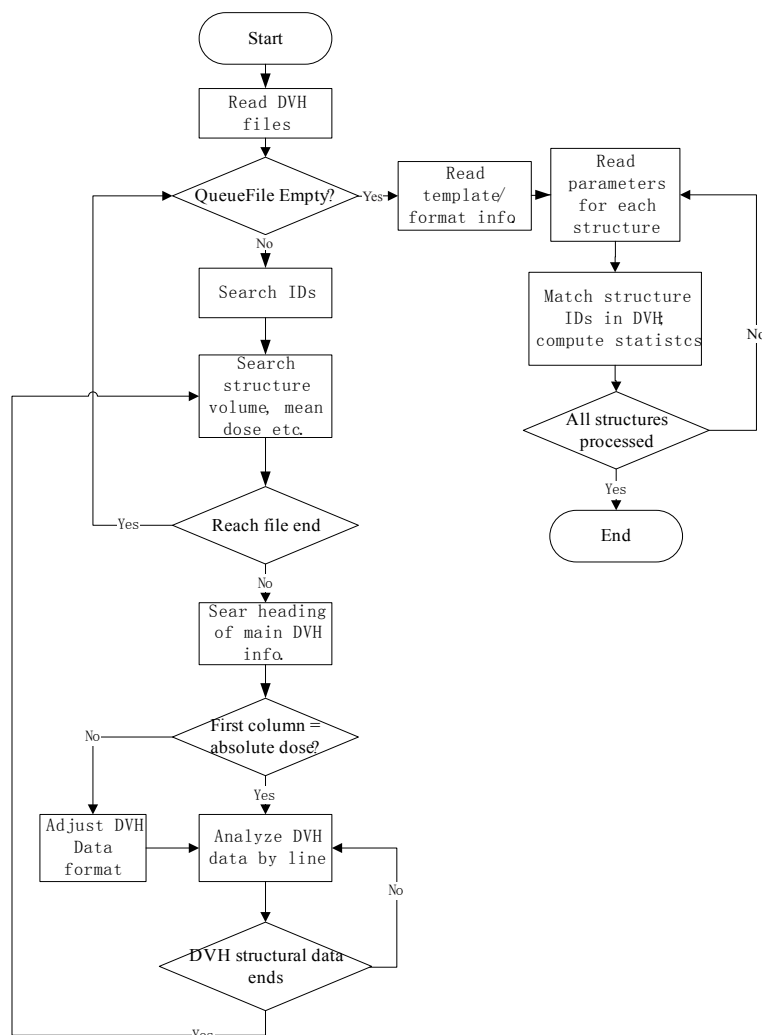


图2 DVH数据处理算法

Fig.2 Algorithm for DVH data processing

验证过程表明,基于上述架构和算法的Web工具可以成功识别并自动处理所有DVH表单文件,包括不同顺序的DVH主数据结构和同一结构的多种命名规则等。从打开计划并导出30例DVH数据,到一键上传(20 M带宽,文件总大小180 M)并生成Excel统计结果,基于Web的自动提取过程共耗时约20 min,平均每个病例不到1 min。而从打开计划开始,人工提取整套相同数据大约需要4 h,平均每个病例耗时8 min。通过比较两种方法提取的数据,并对差异较大者逐一进行人工检查,共发现明显错误的人工提取结果15处,涉及12个病例。对修正后的人工提取数据和Web自动提取结果进行统计比对分析,差异结果如表1所示。

### 3 讨论

基于Web的自动提取工具不仅在效率上远胜于人工提取方法,而且避免了人工提取过程可能存在的失误(具体速度和失误率虽然因人而异,但难以完

全避免),从而节约了复核数据的时间和精力。自动提取工具的这些优势还将随着统计病例量的增加将得到进一步显现。

表1数据说明,自动读取和人工分析得出的 $HI_{PTV}$ 、 $D_{50\%_{FH}}$ 、 $D_{50\%_{UB}}$ 、 $D_{mean_{FH}}$ 、 $D_{mean_{UB}}$ 的差异极小且可以忽略( $\leq 0.01$ ,  $P > 0.05$ ),其主要来源是肉眼观察和测量所产生的误差以及精确的小数位数等。值得注意的是,PTV和PGTV的CI值均存在较大差异,分别达到 $(2.60 \pm 1.04)\%$ 和 $(0.66 \pm 0.29)\%$ ,且均具有统计学意义( $P < 0.001$ )。究其原因,人工和程序读取的靶区体积完全一致,故差异全部来自计算CI时所需的 $V_{100\%}$ ,人工提取采用的方法是右键功能中等剂量线结构转换及体积测量工具,而自动读取程序所依据的是DVH数据中“body”结构100%处方剂量所对应的体积。由于不同厂家的计划系统之间或同一计划系统的不同分析模块之间可能采用不同的空间采样算法(Spatial Sampling Algorithm),导致位于结构边界上的部分体素会因自身的体积以及不同的统计标准而

表1 基于Web自动提取与人工分析(参考)的30例DVH数据差异统计分析

Tab.1 Statistical comparison of differences between Web-based automated extraction and manual analysis (reference) for 30 DVH

Parameter	Mean	95% Confidence interval		Standard deviation	P value
		Lower	Upper		
HI <sub>PGTV</sub>	0.03	0.01	0.06	0.066	0.014
CI <sub>PGTV</sub>	2.60	2.21	2.98	1.035	<0.001
V <sub>107%_PGTV</sub>	-0.45	-1.32	0.42	2.342	0.465
HI <sub>PTV</sub>	-0.01	-0.02	0	0.032	0.215
CI <sub>PTV</sub>	0.66	0.55	0.77	0.294	<0.001
D <sub>50%_FH</sub>	0	0	0.01	0.016	0.781
D <sub>mean__FH</sub>	0	-0.04	0.03	0.088	0.839
D <sub>50%_UB</sub>	0	-0.01	0	0.017	0.416
D <sub>mean_UB</sub>	-0.01	-0.02	0.01	0.051	0.485

Differences were expressed as: (auto-manual)/manual\*100%.

PGTV: Planning gross target volume; PTV: Planning target volume; HI: Homogeneity index; CI: Conformity index;

FH: Femoral head; UB: Urinary bladder

被纳入或排除。理论上说,等剂量线生成的结构体积越大、边界剂量梯度越小,各种算法之间的一致性就越小<sup>[17]</sup>。这也很好地解释了为何两种方法计算的CI在PTV中的差异远小于PGTV(0.66% vs 2.60%),后者体积小,且边界的剂量跌落更加明显。

虽然Eclipse系统自带计算CI的功能,但一次只允许指定一个目标靶区和处方剂量,并且只能计算该目标靶区的CI值。而对于同步推量计划,如果要计算另一靶区的CI值,则需要同时修改目标靶区和处方剂量。但如果原计划曾以PGTV或PTV进行过归一运算,则该修改势必导致整个计划的改变,从而丧失可比性。因此,我们仅对比了Eclipse给出的部分CI值,发现其与基于DVH数据独立计算的CI值更加接近,且采用该方法的自动提取工具很好地避免了一次只能计算1个CI值的缺陷。

作为一个初步的可行性探索研究,本工作还有许多需要进一步改进和完善的方面,包括:(1)增加压缩包处理功能,以进一步缩短数据上传时间(已实现,但尚需进一步调试验证);(2)与计划系统厂家协商开放必要端口,实现DVH表单数据的自动下载,从而避免手工导出DVH的工序,提高效率;(3)虽然该工具暂时受到各医院不同的结构命名规则及解剖部位等限制(可依需求在程序中增加内容),但本文提出的DVH数据搜索和自动提取算法本身可以通用,且基于Web的设计有效避免了平台兼容和软件安装

等技术问题。

## 4 结论

本工作开发的基于Web的DVH数据自动提取工具拥有高效、准确、跨平台等特点,可以代替人工测量用于海量病例分析和大数据挖掘,为临床和科研提供便利。

## 【参考文献】

- [1] DRZYMALA R E, MOHAN R, BREWSTER L, et al. Dose-volume histograms[J]. Int J Radiat Oncol Biol Phys, 1991, 21(1): 71-78.
- [2] XING L, LI J G, PUGACHEV A, et al. Estimation theory and model parameter selection for therapeutic treatment plan optimization[J]. Med Phys, 1999, 26(11): 2348-2358.
- [3] CHENG C W, DAS I J. Treatment plan evaluation using dose-volume histogram (DVH) and spatial dose-volume histogram (zDVH)[J]. Int J Radiat Oncol Biol Phys, 1999, 43(5): 1143-1150.
- [4] OETZEL D, SCHRAUBE P, HENSLEY F, et al. Estimation of pneumonitis risk in three-dimensional treatment planning using dose-volume histogram analysis[J]. Int J Radiat Oncol Biol Phys, 1995, 33(2): 455-460.
- [5] WARKENTIN B J, STAVREV P, STAVREVA N, et al. A TCP-NTCP estimation module using DVHs and known radiobiological models and parameter sets[J]. J Appl Clin Med Phys, 2004, 5(1): 50-63.
- [6] ZHEN H, NELMS B E, TOMÉ W A. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA [J]. Med Phys, 2011, 38(10): 5477-5489.

(下转462页)