

基于卷积神经网络的硬膜下和硬膜外血肿分割方法的一致性评价

田风选 常健博 陈亦豪 魏俊吉 冯铭 王任直 贺喜武

【摘要】 **目的** 探讨基于卷积神经网络的血肿分割算法对硬膜下和硬膜外血肿的测量结果与手动分割结果的一致性。**方法** 纳入 2017 年 1 月至 2019 年 6 月中国颅内出血影像数据库 129 例硬膜下和硬膜外血肿患者计 352 张 CT 影像(硬膜下血肿 33 例计 104 张影像、硬膜外血肿 96 例计 248 张影像),均采用手动分割、算法分割、多田公式 3 种方法对血肿体积进行测量,以手动分割作为“金标准”,分别与算法分割和多田公式进行一致性检验,并探讨血肿形态和边界对算法的影响。**结果** 与多田公式相比,算法分割的百分误差最小(23.62%),有 94.89%(334/352)的测值在 95% 一致性界限(95%LoA)内,与“金标准”的一致性良好;但算法分割的波动范围更大,在不对称($P=0.000$)和边界清晰($P=0.000$)的血肿中表现更佳。**结论** 基于卷积神经网络构建的算法分割具有一定的临床应用前景,但尚待进一步验证。

【关键词】 血肿,硬膜下; 血肿,硬膜外,颅内; 人工智能; 神经网络(计算机); 体层摄影术, X 线计算机

Agreement evaluation of an automatic segmentation algorithm for quantifying subdural/epidural hemorrhage volume using convolution neural network

TIAN Feng-xuan¹, CHANG Jian-bo², CHEN Yi-hao², WEI Jun-ji², FENG Ming², WANG Ren-zhi², HE Xi-wu¹
¹Department of Neurosurgery, The Fifth People's Hospital of Qinghai Province, Xining 810007, Qinghai, China

²Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China

Corresponding author: HE Xi-wu (Email: 1684793310@qq.com)

【Abstract】 **Objective** To validate the agreement among the convolution neural network segmentation algorithm, Tada formula and manual segmentation for subdural/epidural hemorrhage volume. **Methods** A total of 129 cases with 352 subdural/epidural hemorrhage CT scans were extracted from Chinese Intracranial Hemorrhage Image Database (CICHID) from January 2017 to June 2019. All CT scans were measured by three methods including manual segmentation, algorithm segmentation and Tada formula. The manual segmentation was regarded as the "golden standard" and the agreement test among three methods was performed. We explored the influence factors in different measurement methods, such as the shape or boundary of hematoma. **Results** Compared with the Tada formula method, the percentage error of segmentation algorithm was small (23.62%), and the agreement between algorithm and the manual reference was strong, which 94.89% (334/352) of the data was within the 95% limits of agreement (95%LoA), however, the 95%LoA was broad. And the performance of segmentation algorithm showed better in asymmetry ($P=0.000$) and clear boundary hematoma ($P=0.000$). **Conclusions** The segmentation algorithm based on convolution neural network has a certain application prospect, but need to be validated in large sample research.

【Key words】 Hematoma, subdural; Hematoma, epidural, cranial; Artificial intelligence; Neural networks (computer); Tomography, X-ray computed

Conflicts of interest: none declared

doi:10.3969/j.issn.1672-6731.2021.03.011

作者单位:810007 西宁,青海省第五人民医院神经外科(田风选,贺喜武);100730 中国医学科学院北京协和医学院北京协和医院神经外科(常健博,陈亦豪,魏俊吉,冯铭,王任直)

通讯作者:贺喜武,Email:1684793310@qq.com

精确计算硬膜下血肿(SDH)和硬膜外血肿(EDH)体积具有重要临床意义^[1]。既往研究发现,对于慢性硬膜下血肿,仅引流 20% 的出血即可有效帮助血肿吸收^[2]。不同类型颅内出血有多种测量和计算方法,传统的手动分割方法系计算血肿体积的“金标准”,但该方法耗时费力;另一传统的方法多田公式,即长×宽×高/2(A×B×C/2),虽适用于脑实质出血、硬膜下血肿^[3]和硬膜外血肿^[4],但因硬膜下和硬膜外血肿形态并非均为对称的新月形或纺锤形,特别是许多慢性硬膜下血肿呈逗号样^[5],故计算结果的准确性有待商榷。随着深度学习等计算机技术的进步,越来越多的学者开始尝试应用相关技术计算血肿体积,如 Cho 等^[6]构建涵盖所有脑出血类型的血肿体积算法,准确度达 80%、回归度为 82%;国内常健博等^[7]也发现,与多田公式等传统算法相比,卷积神经网络(CNN)与手动分割估算脑实质出血血肿体积的一致性更高。但迄今未见一种专门针对硬膜下和硬膜外血肿的算法,基于此,本研究拟提出一种基于卷积神经网络的算法分割(以下简称算法分割),并与多田公式、手动分割进行对比,以期探索一种精准、便捷的硬膜下和硬膜外血肿体积测量方法。

资料与方法

一、研究资料

1. 资料来源 本研究所用影像学资料经中国颅内出血影像数据库(CICHID)授权^[8],该数据库由中国医学科学院北京协和医院牵头,于 2019 年成立,截至 2020 年 6 月,该数据库共纳入 20 余所医疗中心计 5000 余例脑出血患者的头部 CT 影像资料。该数据库中的颅内出血主要以自发性脑实质出血为主,同时还包括脑室出血、硬膜下血肿、硬膜外血肿、蛛网膜下腔出血等类型,本研究选择硬膜下和硬膜外血肿的数据进行分析。

2. 纳入与排除标准 (1)选择硬膜下和硬膜外血肿患者的头部 CT 影像。(2)均为术前或未行手术治疗时获得。(3)均为标准的 DICOM 格式,扫描参数及机器品牌不限。(4)年龄 > 18 岁。(5)排除存在双侧硬膜下或硬膜外血肿的患者。

3. 一般资料 根据上述纳入与排除标准,选择 2017 年 1 月至 2019 年 6 月共 129 例患者计 352 张(包含 10 208 层图片)CT 影像,其中,硬膜下血肿患者有 33 例共 104 张 CT 扫描结果、硬膜外血肿患者 96 例

共 248 张 CT 扫描结果。

二、研究方法

1. 血肿体积估算方法 本研究纳入的 352 张 CT 影像均采用手动分割、算法分割、多田公式 3 种方法对血肿体积进行测量。(1)手动分割:采用 Insight Toolkit SNAP 开源软件(ITK-SNAP, <https://itk.org/>)^[9],在横断面 CT 平扫影像上沿血肿边界进行勾画,输出血肿体积。手动分割时由一位研究员进行标注,由另一位研究员进行核实,标注边界有争议时与第三者讨论后决定。(2)多田公式:A×B×C/2是在 ITK-SNAP 开源软件辅助下,精确测量最大血肿面积层面最长径(A),以及该层面与最长径垂直的最长径(B),同时以实际层厚乘以出血层面数(C)^[10]。由两位研究员分别独立完成 A、B、C 的精确测量,取平均值纳入最终统计。(3)算法分割:首先,为降低构建算法分割结果的过拟合风险,提高鲁棒性,对所纳入病例的 CT 影像进行五折交叉验证。即由计算机将 CT 影像随机均分为 5 组,每次选择 4 组作为训练集,余 1 组作为验证集,共进行 5 次训练,汇总 5 次模型在各自验证集的结果,以评估算法模型的性能。然后,将血肿手动分割数据作为卷积神经网络的输入数据,通过 TensorFlow 平台(<https://tensorflow.google.cn>)搭建三维卷积神经网络^[9]。网络结构为 3D U-Net,选择自适应矩估计(Adam)优化器作为迭代过程中的优化器,以权重交叉熵作为损失函数进行计算,以手动分割结果作为“金标准”,采用 Dice 相似性系数(DSC)作为分割精度标准进行训练。训练过程中实时检查损失函数和验证集上的分割精度,对收敛速度和趋势进行判断并调整网络参数,于训练集获得满意模型,在验证集获得算法分割结果。

2. 血肿形态及边界对测量方法的影响 为进一步分析不同测量方法一致性检验存在差异的原因,本研究对不同测量方法的百分误差在不同血肿形态(对称与不对称)、血肿边界(模糊与清晰)中是否存在差异进行分析。其中,纺锤形血肿标记为对称,凹透镜形、逗号形等血肿标记为不对称;由标注员凭借经验根据血肿边界与脑组织之间 CT 值差距的大小判断血肿边界清晰或模糊。

3. 统计分析方法 采用绝对差值和百分误差评估不同算法的准确性,百分误差的具体计算方法为 $|V_{其他方法} - V_{手动分割}| / V_{手动分割} \times 100\%$;一致性检验采用 Bland-Altman 分析(均数 - 差值)、组内相关

系数 (ICC) 表示, 并采用 PRISM GraphPad 8.0 软件绘制 Bland-Altman 分析图。采用 SPSS 20.0 统计软件进行数据处理与分析, 正态性检验采用 Kolmogorov-Smirnov 检验, 呈非正态分布的计量资料以中位数和四分位数间距 [$M(P_{25}, P_{75})$] 表示, 采用 Mann-Whitney U 检验。以 $P \leq 0.05$ 为差异具有统计学意义。

结 果

经 Kolmogorov-Smirnov 检验, 通过不同算法获得的血肿体积以及各种方法与手动分割的体积百分误差均为非正态分布资料 ($P = 0.000$), 不同算法所得血肿体积和各种方法与手动分割的体积百分误差如表 1 所示, 以手动分割结果为“金标准”, 算法分割的百分误差小于多田公式, 提示算法分割所得血肿体积更接近手动分割。

算法分割对手动分割差值的中位数 (-4.51 ml) 较多田公式对手动分割差值的中位数 (8.39 ml) 小, 提示算法分割结果较多田公式更接近“金标准”。一致性分析显示, 算法分割对手动分割差值的 95% 一致性界限 (95%LoA) 为 -59.47 ~ 42.76 ml, 差值范围为 102.23 ml, 而多田公式对手动分割差值的 95%LoA 为 -15.15 ~ 37.50 ml, 差值范围为 52.65 ml, 提示多田公式的波动范围较小; 两种方法与手动分割的组内相关系数均较高 (> 0.75), 提示两种算法与“金标准”均具有较好的一致性 (表 2, 图 1)。

不同血肿形态和边界在同一种血肿体积测量方法中的差异如表 3 所示, 血肿对称与不对称之间算法分割的百分误差差异有统计学意义 ($P = 0.000$), 提示算法分割在不对称血肿的百分误差更小; 血肿边界 (模糊与清晰) 在算法分割和多田公式的百分误差均有统计学意义 ($P = 0.000$), 其中算法分割在边界清晰的血肿百分误差更小, 而多田公式在边界模糊的血肿中的百分误差更小。

讨 论

本研究采用基于卷积神经网络的颅内血肿自动分割算法, 可自动识别、分割头部 CT 平扫显示的硬膜下和硬膜外血肿, 并计算血肿体积, 以手动分割为“金标准”, 分别对算法分割、多田公式进行一致性检验, 结果显示, 与多田公式相比, 算法分割的百分误差最小, 为 23.62%; Bland-Altman 分析显示, 有 94.89% (334/352) 的测值在 95%LoA 内, 与“金标

表 1 不同算法测量的硬膜下和硬膜外血肿体积和百分误差 [$M(P_{25}, P_{75})$]

Table 1. Subdural/epidural hemorrhage volumes and percentage error by different measurement methods [$M(P_{25}, P_{75})$]

算法	例数	血肿体积 (ml)	百分误差 (%)
手动分割	352	35.52 (15.73, 56.46)	—
算法分割	352	29.05 (11.17, 49.06)	23.62 (10.96, 56.29)
多田公式	352	48.75 (19.84, 74.25)	28.20 (12.28, 45.93)

—, not available, 不适用

准”的一致性良好; 但算法分割的 95%LoA 波动范围更大, Mann-Whitney U 检验显示血肿形态和边界对算法分割的精确性有影响, 算法分割测量形态不对称、边界清晰的血肿可能更佳。

多田公式的计算方法是由椭球体体积公式 ($4/3 \times \pi \times 0.5$ 长 $\times 0.5$ 宽 $\times 0.5$ 高) 简化而来, 作为床旁简便、快捷的血肿体积估算工具, 广泛应用于脑实质出血血肿的测量。也有学者探索其在硬膜下和硬膜外血肿中的应用。Petersen 和 Espersen^[4] 较早证实该公式可用于硬膜外血肿体积的估算, 其平均误差为 -2.7 ml (标准差为 3.7 ml), 但该项研究纳入标准严格限制为纺锤形硬膜外血肿, 血肿量较小, 故临床推广性较差。Gebel 等^[11] 发现, 多田公式与手动分割算法在急性硬膜下血肿中的相关性较高 ($r = 0.842$), 但相关性低于脑实质出血 ($r = 0.929$)。Sucu 等^[1] 验证多田公式在慢性硬膜下血肿中的应用, 发现长径 A 和宽径 B 在血肿面积最大层面进行测量, 优于在中间层面。但上述研究均存在一定局限性, 首先是样本量较小, 均 < 50 例; 其次, 比较的是不同算法对同一整体的测量结果, 但统计学指标并未采用一致性评价而采用相关系数 (correlation coefficients)^[12]; 最后, 多田公式在慢性硬膜下血肿中高估血肿体积约 10%^[5]。与上述研究结果类似, 本研究结果显示, 多田公式可用于硬膜下和硬膜外血肿测量, 其组内相关系数较高 (0.974)。但本研究多田公式的百分误差较既往研究偏大 (28.20%), 平均高估出血量约 8 ml, 究其原因, 可能是由于多田公式的误差随血肿体积的增加而增大^[13]。

为弥补传统算法的不足, 近年越来越多的研究应用神经网络相关算法用于血肿分割^[6,14], 其准确度 (Dice 值) 高达 80%^[6]。本研究算法分割的百分误差较多田公式小 (23.62% 对 28.20%), 绝对误差更小

表2 不同测量方法的一致性检验 (n = 352)

Table 2. Agreement test of different measurement methods (n = 352)

血肿差值	差值范围(ml)	均数(ml)	中位数(ml)	95%LoA(ml)	ICC(95%CI)
算法分割对手动分割	-167.53 ~ 178.81	-11.45	-4.51	-59.47 ~ 42.76	0.858(0.825 ~ 0.885)
多田公式对手动分割	-25.38 ~ 77.02	2.22	8.39	-15.15 ~ 37.50	0.974(0.968 ~ 0.979)

LoA, limits of agreement, 一致性界限; ICC, intraclass correlation coefficient, 组内相关系数

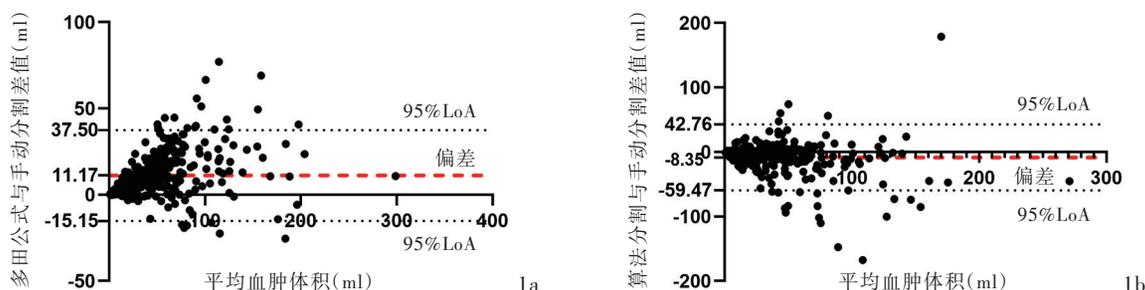


图1 Bland-Altman 一致性检验 1a 多田公式对手动分割的95%LoA范围为-15.55 ~ 37.50 ml, 约94.60% (333/352)的测值在95%LoA内 1b 算法分割对手动分割的95%LoA为-59.47 ~ 42.76 ml, 约94.89% (334/352)的测值在95%LoA内

Figure 1 Bland-Altman agreement test findings The 95%LoA of Tada formula vs. manual segmentation was -15.55-37.50 ml and the points inside 95%LoA was 94.60% (333/352, Panel 1a). The 95%LoA of algorithm segmentation vs. manual segmentation was -59.47-42.76 ml and the points inside 95%LoA was 94.89% (334/352, Panel 1b).

表3 同一算法百分误差在不同血肿形态和边界中的比较 [M(P₂₅, P₇₅), %]

Table 3. Analysis of different hematoma shape and boundary in the same measurement method [M (P₂₅, P₇₅), %]

算法	血肿形态		Z值	P值	血肿边界		Z值	P值
	对称(n = 223)	不对称(n = 129)			模糊(n = 97)	清晰(n = 255)		
算法分割	29.31(13.15, 63.83)	17.37(9.16, 41.40)	-3.659	0.000	42.96(15.54, 72.91)	20.36(9.73, 45.41)	-3.562	0.000
多田公式	30.20(11.00, 52.81)	26.91(15.69, 40.49)	-0.661	0.508	17.71(8.16, 30.97)	33.01(15.91, 52.43)	4.834	0.000

(-4.51 ml 对 8.39 ml)。Bland - Altman 图证实有 94.89% (334/352) 的测值在 95%LoA 内, 显示与“金标准”的一致性良好, 但其波动范围较大 (-59.47 ~ 42.76 ml), 其原因主要是算法分割在多中心数据集上泛化能力较差, 在测试中存在极值案例, 导致约 5.11% (18/352) 在 95%LoA 以外的误差远超过多田公式; 其次, 本研究采用五折交叉验证, 即数据集随机分为 5 组轮流作为训练集和验证集, 最终结果显示 5 种算法模型的平均能力, 导致其 95%LoA 不仅反映算法分割与“金标准”之间的差异, 而且反映不同算法模型之间的差异, 导致其与多田公式相比, 组内相关系数较低。

本研究还发现, 无论是多田公式还是算法分割的准确性, 均受血肿边界清晰度的影响。这可能是由于慢性硬膜下血肿的边界不清, 在手工标注边界、测量径线时存在误差, 即使双人背对背标注, 仍存在较大的标注误差, 难以确定实际出血量; 同时这种标注误差也被引入算法分割训练中, 使得算法

分割的精确性降低。同时算法分割还受血肿形态的影响, 其在不对称血肿中的百分误差优于对称血肿 (17.37% 对 29.31%), 而多田公式无论血肿形态其百分误差相近, 约为 30%, 提示算法分割在形态不规则血肿中存在优势。

为了更好地把握病情、更精准地回答临床问题, 如何简便快捷地计算颅内血肿体积, 特别针对传统多田公式计算不准确的硬膜下和硬膜外血肿, 是临床医师面临的挑战。脑实质出血的血肿体积被作为判断患者是否需行手术治疗的指征之一^[15]; 但在硬膜下和硬膜外血肿中, 由于多田公式存在较大误差, 而手动分割又耗时费力, 限制其在血肿体积与预后关系研究中的应用。简便、准确的硬膜下和硬膜外血肿体积测量工具的开发与应用与患者预后的关系, 使探索血肿体积作为手术指征成为可能。随着人工智能(AI)技术在临床医学领域的深入应用, 逐渐实现自动、精准和便捷的颅内血肿分割, 是未来影像组学、手术模拟和手术导航等算法

进一步研究的基础。

结 论

在硬膜下和硬膜外血肿中,与传统多田公式相比,采用卷积神经网络计算头部 CT 平扫图像中的血肿体积,其百分误差较小,与手动分割结果的一致性良好,在形态不对称、边界清晰的血肿中表现更好;但算法分割在少数病例中的误差较大,临床应用需进一步探索。

利益冲突 无

参 考 文 献

- [1] Sucu HK, Gokmen M, Gelal F. The value of XYZ/2 technique compared with computer-assisted volumetric analysis to estimate the volume of chronic subdural hematoma[J]. Stroke, 2005, 36: 998-1000.
- [2] Tabaddor K, Shulmon K. Definitive treatment of chronic subdural hematoma by twist-drill craniostomy and closed-system drainage[J]. J Neurosurg, 1977, 46:220-226.
- [3] Manickam A, Marshman LAG, Johnston R, Thomas PAW. Mathematical formulae to estimate chronic subdural haematoma volume: flawed assumption regarding ellipsoid morphology[J]. J Clin Neurosci, 2017, 40:39-43.
- [4] Petersen OF, Espersen JO. Extradural hematomas: measurement of size by volume summation on CT scanning [J]. Neuroradiology, 1984, 26:363-367.
- [5] Zhao KJ, Zhang RY, Sun QF, Wang XQ, Gu XY, Qiang Q, Gao C, Shen JK. Comparisons of 2/3Sh estimation technique to computer-assisted planimetric analysis in epidural, subdural and intracerebral hematomas[J]. Neurol Res, 2010, 32:910-917.
- [6] Cho J, Park KS, Karki M, Lee E, Ko S, Kim JK, Lee D, Choe J, Son J, Kim M, Lee S, Lee J, Yoon C, Park S. Improving sensitivity on identification and delineation of intracranial hemorrhage lesion using cascaded deep learning models[J]. J Digit Imaging, 2019, 32:450-461.
- [7] Chang JB, Jiang SZ, Chen XJ, Lok KH, Lee YL, Zhang QH, Wei JJ, Shi L, Feng M, Wang RZ. Consistency evaluation of an automatic segmentation for quantification of intracerebral hemorrhage using convolution neural network [J]. Zhongguo Xian Dai Shen Jing Ji Bing Za Zhi, 2020, 20:585-590.[常健博, 姜崇种, 陈显金, 骆嘉希, 李沃霖, 张庆华, 魏俊吉, 石林, 冯铭, 王任直. 基于卷积神经网络的自发性脑出血血肿分割方法的一致性评价[J]. 中国现代神经疾病杂志, 2020, 20:585-590.]
- [8] Wang RZ, Chang JB, Feng M. Prospects for precious diagnosis, assessment, prediction and treatment of hemorrhagic stroke[J]. Zhongguo Xian Dai Shen Jing Ji Bing Za Zhi, 2019, 19:618-621. [王任直, 常健博, 冯铭. 出血性卒中精准诊断、评估、预测及治疗展望[J]. 中国现代神经疾病杂志, 2019, 19:618-621.]
- [9] Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User - guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability[J]. Neuroimage, 2006, 31:1116-1128.
- [10] Kothari RU, Brott T, Broderick JP, Barsan WG, Sauerbeck LR, Zuccarello M, Khoury J. The ABCs of measuring intracerebral hemorrhage volumes[J]. Stroke, 1996, 27:1304-1305.
- [11] Gebel JM, Sila CA, Sloan MA, Granger CB, Weisenberger JP, Green CL, Topol EJ, Mahaffey KW. Comparison of the ABC/2 estimation technique to computer-assisted volumetric analysis of intraparenchymal and subdural hematomas complicating the GUSTO-1 trial[J]. Stroke, 1998, 29:1799-1801.
- [12] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HC, Lijmer JG; Standards for Reporting of Diagnostic Accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration[J]. Clin Chem, 2003, 49:7-18.
- [13] Wang CW, Juan CJ, Liu YJ, Hsu HH, Liu HS, Chen CY, Hsueh CJ, Lo CP, Kao HW, Huang GS. Volume - dependent overestimation of spontaneous intracerebral hematoma volume by the ABC/2 formula[J]. Acta Radiol, 2009, 50:306-311.
- [14] Chang PD, Kuoy E, Grinband J, Weinberg BD, Thompson M, Homo R, Chen J, Abcede H, Shafie M, Sugrue L, Filippi CG, Su MY, Yu W, Hess C, Chow D. Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT [J]. AJNR Am J Neuroradiol, 2018, 39:1609-1616.
- [15] Gross BA, Jankowitz BT, Friedlander RM. Cerebral intraparenchymal hemorrhage: a review [J]. JAMA, 2019, 321: 1295-1303.

(收稿日期:2021-03-10)

(本文编辑:袁云)

《中国现代神经疾病杂志》关于谨防伪造微信采编中心的声明

《中国现代神经疾病杂志》编辑部近期发现伪造本刊微信采编中心的非法行为,微信号 1025282431,昵称麦芽糖,伪造《中国现代神经疾病杂志》采编中心。该微信号以核对作者信息为由,请我刊作者添加其为微信好友,借以窃取相关信息甚至索取审稿费和版面费等,此举对我刊及广大作者、读者造成严重不良影响。

《中国现代神经疾病杂志》特此郑重声明:我刊迄今为止并未建立微信平台的采编中心,作者投稿的唯一途径是登录我刊官方网站 www.xdjb.org,进入“作者在线投稿”界面,按照操作提示提交稿件。稿件经外审通过后,需作者配合修改,达到发表要求后方可待编、排期和刊出,这一过程中编辑部人员与作者之间的联系均采用我刊公共邮箱(xdsjbbz@263.net.cn)和公用电话[(022)59065611,59065612]。

若遇假冒我刊网站、伪造我刊采编中心、中介、代理等不法事件,欢迎广大作者和读者向我刊提供相关线索!对于以我刊名义从事非法活动的个别网站或微信号,我刊保留通过法律途径解决问题的权利。此声明长期有效,最终解释权归我刊所有。