

## Progresses of deep learning in temporal bone CT structures segmentation

YOU Zixuan<sup>1</sup>, XUE Zhiyuan<sup>2</sup>, JIN Richu<sup>3</sup>, ZHANG Ke<sup>1\*</sup>

(1. Department of Otolaryngology Head and Neck Surgery, Peking University Third Hospital, Beijing 100191, China;

2. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;

3. Tech X Academy, Shenzhen Polytechnic University, Shenzhen 518055, China)

[Abstract] Accurate segmentation of temporal bone CT structures is a key for ear surgery planning and diagnosis and treatment of diseases, but traditional manual segmentation methods have limits such as long-time consumption and low accuracy. Deep learning (DL) has shown significant advantages in the field of medical imaging, providing a new solution for structures segmentation of temporal bone CT. The progresses of DL in temporal bone CT structures segmentation were reviewed in this article.

[Keywords] temporal bone; tomography, X-ray computed; deep learning; image segmentation

DOI:10.13929/j.issn.1003-3289.2026.03.028

## 深度学习用于分割颞骨 CT 结构进展

游子轩<sup>1</sup>, 薛智元<sup>2</sup>, 金日初<sup>3</sup>, 张珂<sup>1\*</sup>

(1. 北京大学第三医院耳鼻咽喉头颈外科, 北京 100191; 2. 中国科学院空天信息创新研究院,

北京 100094; 3. 深圳职业技术大学未来技术学院, 广东 深圳 518055)

[摘要] 精准分割颞骨 CT 结构是耳科手术规划与诊疗疾病的关键, 而传统手动分割存在耗时长、精度低等问题。深度学习技术在医学影像领域展现出显著优势, 为分割颞骨 CT 结构提供了新的解决方案。本文围绕深度学习用于颞骨 CT 结构分割进展进行综述。

[关键词] 颞骨; 体层摄影术, X 线计算机; 深度学习; 图像分割

[中图分类号] R762; R814.42 [文献标识码] A [文章编号] 1003-3289(2026)03-0457-05

颞骨内含听觉与平衡觉的关键解剖结构, 如内耳迷路和听骨链, 形态复杂、结构细微。手动分割颞骨 CT 结构较为耗时且戴斯相似系数 (Dice similarity coefficient, DSC) 欠稳定<sup>[1]</sup>, 难以满足临床需求。近年来, 深度学习 (deep learning, DL) 凭借端到端的特征学习及三维上下文建模特点, 用于分割脑白质<sup>[2]</sup>、冠状动脉<sup>[3]</sup>和胰腺<sup>[4]</sup>等表现出色, 但用于分割颞骨 CT 结构仍面临标注样本稀缺、结构尺度差异大及个体解剖变异等挑战。研究<sup>[5]</sup>显示, 基于传统 3D U-Net 自动分割迷路、

听骨链及面神经管的 DSC 仅为 0.79、0.64 及 0.49。本文围绕 DL 用于分割颞骨 CT 结构进展进行综述。

### 1 算法创新

U-Net 是用于分割医学图像任务的最经典的 DL 架构 (图 1)<sup>[6]</sup>, 可通过跳跃连接联系编码器和解码器, 针对小样本医学数据而实现高效分割。针对颞骨结构复杂、尺度差异大等特点, 研究者<sup>[7-16]</sup>在 U-Net 架构基础上提出多结构协同分割、分割单一精细结构和轻量化架构设计等诸多创新方法。

[基金项目] 国家自然科学基金 (62271008)。

[第一作者] 游子轩 (2003—), 男 (土家族), 湖南常德人, 在读本科。研究方向: 颞骨 CT 图像处理。E-mail: 2210117128@stu.pku.edu.cn

[通信作者] 张珂, 北京大学第三医院耳鼻咽喉头颈外科, 100191。E-mail: kezhang\_ent@bjmu.edu.cn

[收稿日期] 2025-07-14 [修回日期] 2025-11-09

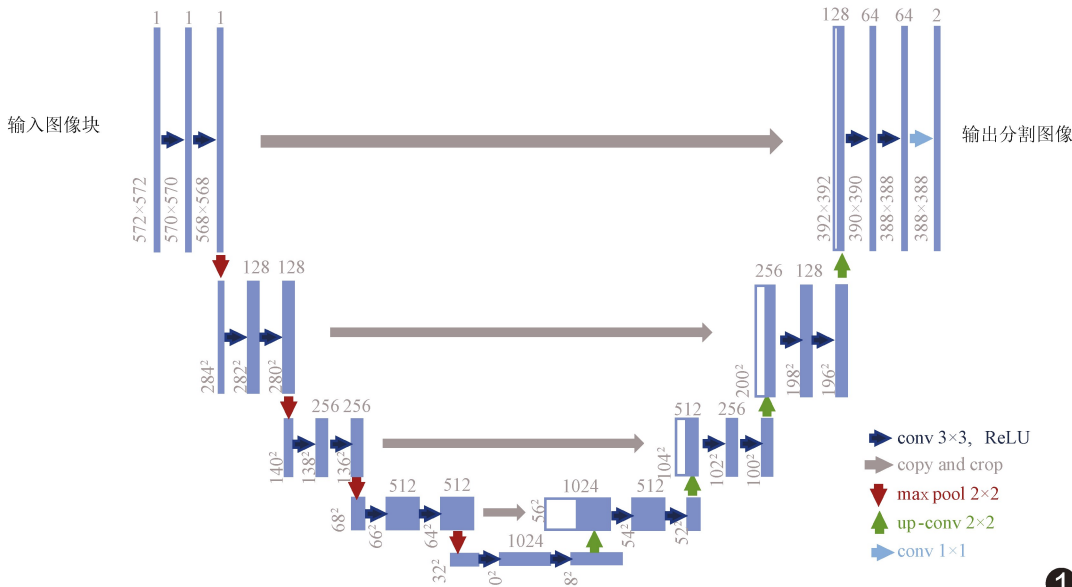


图 1 U-Net 框架

1.1 多结构协同分割 多结构协同分割是实现高精度分割的关键方向之一。LI 等<sup>[7]</sup>在 3D U-Net 基础上改进编码模块,引入深度监督机制而构建的 3D 深度监督密集网络(3D deeply supervised dense network, 3D DSD-Net)在一定程度上提升了分割小器官的精度,其自动分割颞骨 9 个关键结构的 DSC 为 0.77。NIKAN 等<sup>[8]</sup>在 DenseVNet 基础上引入平衡加权采样策略而构建了块状密集连接 3D 网络(patch-wise densely connected-3D net, PWD-3D Net,图 2)能兼顾计算效率和保留小结构特征,其同步分割颞骨 8 个结构的整体 DSC 达 0.86。FAUSER 等<sup>[9]</sup>在具有概率主动形状模型的 3D U-Net 中引入形状正则化,并通过贝塞尔样条和随机树优化而有效减少了过分割及欠分割现象,使分割耳蜗的 DSC 从 0.82 提升至 0.87。

1.2 分割单一精细结构 整体分割与精细结构分割相辅相成。LI 等<sup>[10]</sup>构建的三维多特征融合网络(3D-MFF Net)通过池化特征融合和膨胀卷积模块而实现精准分割外侧半规管(DSC 为 0.73)。WU 等<sup>[11]</sup>在 3D U-Net 架构

中加入空间与通道注意力机制,其分割半规管的 DSC 达 0.93,但用于分割儿童颞骨 CT 时出现效能下降。HUSSAIN 等<sup>[12]</sup>提出基于自适应上下文的级联 2D U-Net 结构 AutoCasU-Net,结合 3D 细化算法可快速精确分割内耳结构(整体 DSC 为 0.90)。

① NEVES 等<sup>[13]</sup>开发的端到端自动分割

耳蜗流程的 DSC 达 0.92,结合 Unity3D 平台进行三维建模可将测量耳蜗导管长度误差控制在 0.57 mm 内。HEUTINK 等<sup>[14]</sup>基于高分辨率 CT (high resolution CT, HRCT)、融合多尺度残差卷积神经网络(convolutional neural network, CNN)与形态学细化算法而实现高精度分割耳蜗(DSC 为 0.90),其体积误差仅 8.4%,具备良好术前评估潜力。

1.3 轻量化架构设计 为满足临床实时处理的需求,轻量化网络设计成为分割颞骨 CT 相关研究的重要方向。LV 等<sup>[15]</sup>提出轻量级三维 CNN W-Net,利用浅层卷积和跳跃连接减少特征信息丢失,并结合形态学增强和区域生长优化分割结果,其参数为传统 3D U-Net 的十分之一而精度与之相当,用于分割耳蜗、听骨链和面神经的 DSC 分别为 0.90、0.85 及 0.77。在此基础上,WANG 等<sup>[16]</sup>进一步研究发现 W-Net 用于分割畸形耳蜗及断裂听骨链亦可保持较高精度(DSC 分别为 0.78 和 0.70),但分割异常面神经的效能有所下降(DSC 仅 0.51),表明模型具有一定泛化性能,但仍需优化。

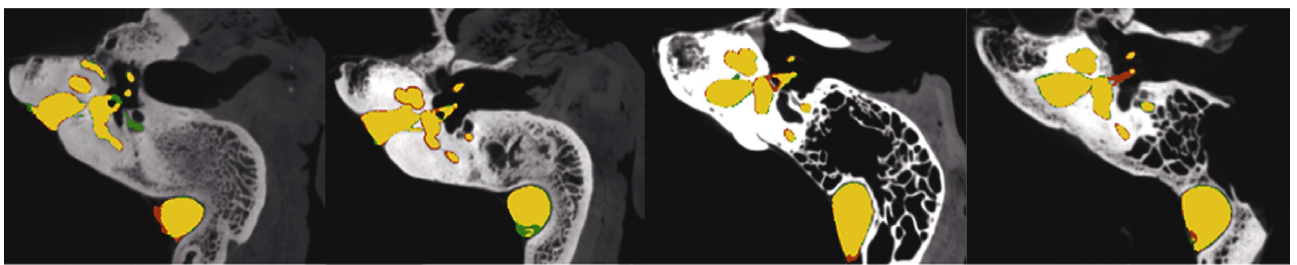


图 2 基于 PWD-3DNet 模型自动分割颞骨矢状位 CT<sup>[8]</sup> (绿色代表真实值,红色代表预测分割标签,黄色代表二者之间重叠区域)

1.4 现有模型效能对比 现有 DL 模型分割颞骨 CT 结构的精度较传统 3D U-Net 显著提升(表 1),但识别面神经、镫骨等微小结构的能力仍有不足。需要注意的是,各项研究所用数据集和评价指标均存在差异,对比结果仅具参考意义;未来应在提高 DL 分割精细结构能力的同时完善处理效率。

## 2 数据优化

DL 模型训练效果高度依赖大规模、高质量的标注数据。颞骨 CT 数据获取困难,病理解剖变异样本稀缺<sup>[23]</sup>;且颞骨结构复杂,内含面神经、听小骨等精细结构,对标注精度提出更高要求(图 3)。

2.1 数据增强 为缓解训练数据不足及类别不平衡问题,分割颞骨影像时常采用数据增强策略,即通过适当修改训练集数据生成新样本,以提升模型泛化能力。LI 等<sup>[23]</sup>提出子特征正交解耦方法,将深度特征分解为类相关与类无关子空间,通过重组生成解剖合理的异常样本;其利用此法检测颈静脉球骨壁缺失的准确率达 96.27%,为精细分割颞骨结构提供了新的手段。

2.2 主动学习 标注医学图像成本较高,主动学习可自动筛选高信息量的样本供专家标注,进而显著降低人工成本。传统的基于不确定性采样方法,如蒙特卡洛 Dropout(Monte Carlo Dropout, MC Dropout)和最大熵采样均依赖批次选择,且易偏重极端样本而忽略整体分布<sup>[24]</sup>。LI 等<sup>[25]</sup>提出的 Labyrinth Net 可结合多模型不一致性框架和观察者网络对筛选样本进行可靠性评估,并以最大连通概率图(maximum connectivity probability map, MCP-Map)消除异常值,最终仅利用 40%的标注数据即可实现自动分割内耳迷路(DSC 为 0.95),与全量标注训练所获结果(DSC 为 0.96)相当;且未来有望扩展至分割听小骨等微小结构。

表 1 现有 DL 模型分割颞骨 CT 效能

模型	分割目标	DSC
3D DSD-Net <sup>[7]</sup>	多结构	锤骨 0.82、砧骨 0.82、外半规管 0.70、整体 0.77
PWD-3D Net <sup>[8]</sup>	多结构	内耳 0.90、乙状窦 0.86、面神经 0.74、整体 0.86
W-Net <sup>[15]</sup>	多结构	耳蜗 0.90、听小骨 0.85、面神经 0.77
SwinUNETR <sup>[17]</sup>	多结构	内耳 0.95、听小骨 0.88、面神经 0.83、整体 0.87
AH-Net <sup>[18]</sup>	多结构	内耳 0.91、听小骨 0.85、面神经 0.75、乙状窦 0.86
nnU-Net <sup>[19]</sup>	多结构	锤骨 0.91、镫骨 0.56、面神经 0.86、骨迷路 0.95
3D-MFF Net <sup>[10]</sup>	外半规管	0.73
AutoCasU-Net <sup>[12]</sup>	内耳	0.90
SegResNet <sup>[20]</sup>	耳蜗	0.94*
UNETR <sup>[21]</sup>	耳蜗	正常耳蜗 0.92、畸形耳蜗 0.93
3D V-Net <sup>[22]</sup>	听小骨	锤骨 0.92、镫骨 0.84、砧骨 0.93

注: \*: 分割正常与畸形耳蜗的综合 DSC 值

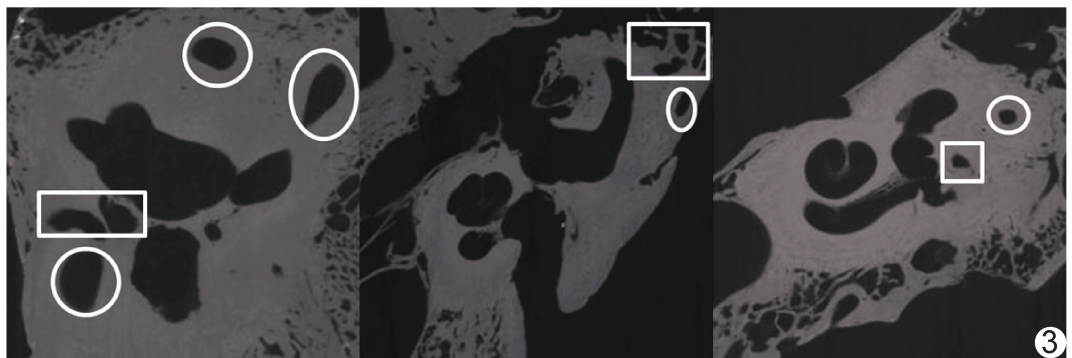


图 3 在颞骨 CT 中标注内耳结构困难<sup>[12]</sup> (圈内为内耳结构、矩形内不属于内耳结构)

## 3 临床应用

3.1 辅助诊断 近年 DL 已广泛用于辅助诊断耳科疾病。DUAN 等<sup>[26]</sup>采用 GoogleNet 和 ResNet 等 CNN 模型鉴别原发性纤毛运动障碍相关中耳炎与单纯性中耳炎伴积液的准确率达 99%,表明 DL 能有效识别影像学特征相似但病理机制不同的耳部疾病,为早期干预提供依据。赵星等<sup>[27]</sup>基于 DL 联合 CT 影像组学自动判别内耳正常和畸形结构的准确率达 96.4%,但因人工标注耗时,目前仅用于小样本数据集。CAI 等<sup>[28]</sup>采用 CNN-UNet 自动分割颞骨 HRCT 中的听小骨并检测其骨折,准确率达 96.2%,展现出潜在辅助诊断价值。CHEN 等<sup>[29]</sup>以 3D CNN 自动分割中耳区域,进而评估慢性中耳炎及胆脂瘤,准确率达 96%,并能生成梯度加权类激活映射(gradient-weighted class activation mapping, Grad-CAM)热图辅助医师理解模型决策,增强了模型可解释性。

3.2 手术规划导航 DL 在耳科手术规划导航中展现出重要价值。TALEB 等<sup>[30]</sup>利用 U-Net 提取儿科手术前 CT 和内镜图像中的鼓膜及锤骨柄解剖轮廓,结合迭代最近点算法实现全自动二维图像配准;据此分割砧骨等关键结构的平均豪斯多夫距离(Hausdorff

distance, HD)为 $(0.98 \pm 0.60)$  mm,整体配准时间不超过 18 s,可满足耳科手术精度和时效性要求。LI 等<sup>[31]</sup>提出的微焦点源、光子计数探测器(photon counting detector, PCD)和 DL 集成的显微 CT 系统能在颞骨常规 200  $\mu$ m CT 扫描基础上进行局部 50  $\mu$ m 的高分辨率成像,通过背景补偿算法与 DL 优化重建实现低剂量高质量成像,具备用于内耳成像及人工耳蜗植入术的潜力。

#### 4 小结与展望

近年 DL 在自动分割颞骨 CT 结构及相关临床应用取得显著进展,分割精度和处理效率均稳步提升,但目前仍存在诸多挑战,如公开可用的标注数据集较少且结构有限、难以满足复杂解剖结构分割需求,以及模型评估体系尚不完善、缺乏统一的标准化效率评估指标及性能评价标准等。未来可进一步聚焦于实时轻量化分割、增强模型鲁棒性和提升泛化性能等并构建完善的综合评估体系,为其临床应用提供有力支持。

利益冲突:全体作者声明无利益冲突。

作者贡献:游子轩查阅文献、撰写和修改文章;薛智元和金日初审阅文章;张珂指导、审阅文章、经费支持。

#### [参考文献]

- [1] LEE J W, ANDERSEN S A W, HITTLE B, et al. Variability in manual segmentation of temporal bone structures in cone beam CT images[J]. *Otol Neurotol*, 2024, 45(3):e137-e141.
- [2] 张伯泉, 麦海鹏, 陈嘉敏, 等. 基于高灰度值注意力机制的脑白质高信号分割[J]. *计算机与现代化*, 2023, 39(12):67-75.
- [3] 潘丹, 骆根强, 曾安. 基于 Transformer 和卷积神经网络双并行分支编码器神经网络的冠状动脉分割[J]. *生物医学工程学杂志*, 2024, 41(6):1195-1203.
- [4] 张文瀚, 王永雄, 曾福斌, 等. 基于前馈注意力 ConvNeXt 模型分割腹部 CT 中的胰腺[J]. *中国医学影像技术*, 2025, 41(3):466-472.
- [5] 柯嘉, 吕弈, 杜雅丽, 等. 三维卷积神经网络在中耳手术相关颞骨影像自动分割中的应用探索[J]. *临床耳鼻咽喉头颈外科杂志*, 2020, 34(10):870-873.
- [6] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]// *International Conference on Medical Image Computing and Computer-assisted Intervention*, 2015. Cham: Springer, 2015: 234-241.
- [7] LI X, GONG Z, YIN H, et al. A 3D deep supervised densely network for small organs of human temporal bone segmentation in CT images[J]. *Neural Netw*, 2020, 124:75-85.
- [8] NIKAN S, van OSCH K, BARTLING M, et al. PWD-3DNet: A deep learning-based fully-automated segmentation of multiple structures on temporal bone CT scans[J]. *IEEE Trans Image Process*, 2021, 30:739-753.
- [9] FAUSER J, BOHLENDER S, STENINI I, et al. Retrospective in silico evaluation of optimized preoperative planning for temporal bone surgery[J]. *Int J Comput Assist Radiol Surg*, 2020, 15(11):1825-1833.
- [10] LI X, FU P, YIN H, et al. A geometric alignment for human temporal bone CT images via lateral semicircular canals segmentation[J]. *Med Phys*, 2022, 49(10):6439-6450.
- [11] WU H, LIU J, CHEN G, et al. Automatic semicircular canal segmentation of CT volumes using improved 3D U-Net with attention mechanism [J]. *Comput Intell Neurosci*, 2021, 2021:9654059.
- [12] HUSSAIN R, LALANDE A, GIRUM K B, et al. Automatic segmentation of inner ear on CT-scan using auto-context convolutional neural network[J]. *Sci Rep*, 2021, 11(1):4406.
- [13] NEVES C A, TRAN E D, COOPERMAN S P, et al. Fully automated measurement of cochlear duct length from clinical temporal bone computed tomography[J]. *Laryngoscope*, 2022, 132(2):449-458.
- [14] HEUTINK F, KOCH V, VERBIST B, et al. Multi-scale deep learning framework for cochlea localization, segmentation and analysis on clinical ultra-high-resolution CT images[J]. *Comput Methods Programs Biomed*, 2020, 191:105387.
- [15] LV Y, KE J, XU Y, et al. Automatic segmentation of temporal bone structures from clinical conventional CT using a CNN approach[J]. *Int J Med Robot*, 2021, 17(2):e2229.
- [16] WANG J, LV Y, WANG J, et al. Fully automated segmentation in temporal bone CT with neural network: A preliminary assessment study[J]. *BMC Med Imaging*, 2021, 21(1):166.
- [17] NEVES C A, CHEMALY T E, FU F, et al. Deep learning method for rapid simultaneous multistructure temporal bone segmentation[J]. *Otolaryngol Head Neck Surg*, 2024, 170(6):1570-1580.
- [18] NEVES C A, TRAN E D, KESSLER I M, et al. Fully automated preoperative segmentation of temporal bone structures from clinical CT scans[J]. *Sci Rep*, 2021, 11(1):116.
- [19] DING A S, LU A, LI Z, et al. A self-configuring deep learning network for segmentation of temporal bone anatomy in cone-beam CT imaging[J]. *Otolaryngol Head Neck Surg*, 2023, 169(4):988-998.
- [20] LI Z, ZHOU L, TAN S, et al. Application of UNETR for automatic cochlear segmentation in temporal bone CTs [J]. *Auris Nasus Larynx*, 2023, 50(2):212-217.
- [21] LI Z, ZHOU L, TAN S, et al. Utilizing deep learning for automatic segmentation of the cochleae in temporal bone computed tomography[J]. *Acta Radiol*, 2025, 66(3):305-311.
- [22] WANG X R, MA X, JIN L X, et al. Application value of a deep

- learning method based on a 3D V-Net convolutional neural network in the recognition and segmentation of the auditory ossicles[J]. *Front Neuroinform*, 2022,16:937891.
- [23] LI X, ZHOU Y, YIN H, et al. Sub-features orthogonal decoupling: Detecting bone wall absence via a small number of abnormal examples for temporal CT images [J]. *Comput Med Imaging Graph*, 2024,115:102380.
- [24] GAILLOCHET M, DESROSIERS C, LOMBAERT H. Active learning for medical image segmentation with stochastic batches[J]. *Med Image Anal*, 2023,90:102958.
- [25] LI X, ZHU Z, YIN H, et al. Labyrinth Net: A robust segmentation method for inner ear labyrinth in CT images [J]. *Comput Biol Med*, 2022,146:105630.
- [26] DUAN B, GUO Z, PAN L, et al. Temporal bone CT-based deep learning models for differential diagnosis of primary ciliary dyskinesia related otitis media and simple otitis media with effusion[J]. *Am J Transl Res*, 2022,14(7):4728-4735.
- [27] 赵星,李晓鸽,高坤,等.基于颞骨 CT 的深度迁移学习放射组学模型辅助鉴别内耳畸形[J]. *临床耳鼻咽喉头颈外科杂志*, 2024, 38(6):547-552.
- [28] CAI Q, ZHANG P, XIE F, et al. Clinical application of high-resolution spiral CT scanning in the diagnosis of auriculotemporal and ossicle[J]. *BMC Med Imaging*, 2024, 24(1):102.
- [29] CHEN B, LI Y, SUN Y, et al. A 3D and explainable artificial intelligence model for evaluation of chronic otitis media based on temporal bone computed tomography: Model development, validation, and clinical application [J]. *J Med Internet Res*, 2024,26:e51706.
- [30] TALEB A, LECLERC S, HUSSEIN R, et al. Registration of preoperative temporal bone CT-scan to otoendoscopic video for augmented-reality based on convolutional neural networks [J]. *Eur Arch Otorhinolaryngol*, 2024,281(6):2921-2930.
- [31] LI M, FANG Z, CONG W, et al. Clinical micro-CT empowered by interior tomography, robotic scanning, and deep learning [J]. *IEEE Access*, 2020, 8:229018-229032.

## 中国学术期刊影响因子年报(2025 版) ——《中国医学影像技术》

“中国学术期刊影响因子年报(2025 版)”于 2025 年 11 月由中国科学文献计量评价研究中心发布。《中国医学影像技术》杂志在临床医学综合类的相关数据为:

- 1 影响力指数(CI 值):**224.743**;
- 2 总被引:**4398**;
- 3 影响因子:**1.072**;
- 4 他引总引比:**0.86**;
- 5 基金论文比:**0.52**;
- 6 WEB 即年下载率:**111**。