

DOI: 10.3969/j.issn.1005-8982.2026.16.010
文章编号: 1005-8982 (2026) 16-0063-06

临床研究·论著

基于递阶拓扑模型的斜视手术规划研究*

王洁月, 吴小影

(中南大学湘雅医院 眼科, 湖南 长沙 410008)

摘要: **目的** 探讨基于递阶拓扑架构的机器学习模型在水平斜视个性化手术规划中的应用价值。**方法** 回顾性分析2021年3月—2024年3月中南大学湘雅医院收治的608例行水平斜视手术患者的临床资料。提取年龄、三棱镜度数、眼轴长度等14项常规术前指标, 采用CatBoost集成学习算法构建两阶段递阶拓扑模型。模型第一阶预测双眼4条水平直肌对应的8种手术操作(后徙或缩短)干预选择, 第二阶预测具体手术量。数据集按80%与20%比例划分为训练集与测试集, 采用宏平均曲线下面积(macro-AUC)、平均绝对误差(MAE)、均方根误差(RMSE)及方案完全命中率评估模型性能, 并引入沙普利加和解释(SHAP)法进行特征解释。**结果** 外斜视组与内斜视组年龄、平均眼轴、双眼平均等效球镜比较, 差异均无统计学意义($P>0.05$)。外斜视组与内斜视组三棱镜度数比较, 差异有统计学意义($P<0.05$)。术前三棱镜度数与各主要干预肌肉左眼外直肌后徙术、右眼外直肌后徙术、左眼内直肌缩短术和右眼内直肌缩短术的实际手术量均呈正相关($P<0.05$)。递阶拓扑模型在手术肌肉选择任务中的macro-AUC为0.962; 手术量预测的MAE为0.491 mm, RMSE为1.494 mm。完整手术方案的完全命中率达到64.8% (79/122)。特征重要性分析表明, 等效球镜度数和术前三棱镜度数等是驱动模型规划的最核心特征。**结论** 递阶拓扑机器学习模型能够高精度模拟临床专家的决策思维, 实现较高精度的个体化手术规划, 具有较强的临床决策支持潜力。

关键词: 水平斜视; 机器学习; 手术规划; 递阶拓扑模型; 临床决策支持

中图分类号: R777.4

文献标识码: A

Research on surgical planning for strabismus based on a hierarchical topological model*

Wang Jie-yue, Wu Xiao-ying

(Department of Ophthalmology, Xiangya Hospital, Central South University, Changsha, Hunan 410008, China)

Abstract: **Objective** To explore the application value of a machine learning model based on a hierarchical topological architecture in personalized surgical planning for horizontal strabismus. **Methods** The clinical data of 608 patients who underwent horizontal strabismus surgery at Xiangya Hospital of Central South University from March 2021 to March 2024 were retrospectively analyzed. Fourteen routine preoperative indicators, including age, prism diopter, and axial length, were extracted, and an ensemble learning algorithm was used to construct a two-stage hierarchical topological model. The first stage of the model predicted the intervention choice among 8 surgical operations (recession or resection) corresponding to the 4 horizontal rectus muscles of both eyes, while the second stage predicted the specific surgical dose. The dataset was divided into a training set and a test set at a ratio of 80% to 20%. Model performance was evaluated using the macro-average area under the curve (macro-AUC), mean absolute error (MAE), root mean square error (RMSE), and the exact match rate of the surgical plans. The SHapley

收稿日期: 2026-05-19

* 基金项目: 国家自然科学基金 (82271091)

[通信作者] 吴小影, E-mail: hewuan@163.com

Additive exPlanations (SHAP) method was introduced for feature interpretation. **Results** There were no significant differences between the exotropia and esotropia groups in age, mean axial length, or mean spherical equivalent ($P > 0.05$). However, the prism diopter differed significantly between the two groups ($P < 0.05$). Preoperative prism diopter was positively correlated with the actual surgical doses of the major target muscle interventions, including left lateral rectus recession, right lateral rectus recession, left medial rectus resection, and right medial rectus resection (all $P < 0.05$). The hierarchical topological model achieved a macro-AUC of 0.962 for the surgical muscle selection task. For surgical dose prediction, the MAE and RMSE were 0.491 mm and 1.494 mm, respectively. The exact match rate for the complete surgical plan was 64.8% (79/122). Feature importance analysis showed that spherical equivalent and preoperative prism diopter were among the most important features driving the model's surgical planning decisions. **Conclusions** The hierarchical topological machine learning model can simulate the decision-making thinking of clinical experts with high precision, achieving highly precise personalized surgical planning, and exhibits strong potential for clinical decision support.

Keywords: horizontal strabismus; machine learning; surgical planning; hierarchical topological model; clinical decision support

水平斜视是一种常见的视觉障碍，在全球范围内影响着约 1.9% 的人口，这种视觉问题不仅关系到个人的外观形象，更重要的是，其可能对双眼视觉功能、深度感知及日常生活的多个方面产生深远的影响^[1]。临床干预策略高度依赖于斜视度数：对于小度数斜视，三棱镜递减、压贴等保守治疗可有效缓解短期复视症状^[2]；但当水平偏斜角 ≥ 15 个三棱镜度数时，临床通常建议尽早行手术干预^[3]。《中国儿童斜视围手术期管理专家共识（2025 年）》^[4]明确强调了术前精准评估与个体化手术设计对保障视觉预后的重要地位。手术量的精准设计不仅关乎眼位的机械矫正，更是术后双眼立体视功能修复及重建的前提，而术后眼位是影响立体视功能恢复的最关键因素之一^[5]。

1 个多世纪以来，眼科医师主要依赖基于三棱镜度数的传统经验公式进行初步的手术规划^[6]。然而，该经验法则对患者的高度异质性特征（如年龄、眼轴长度、屈光状态等）整合处理能力有限^[7]。近期的临床对照研究指出，在使用传统手术规划治疗间歇性外斜视时，即使采用规范的眼外直肌后徙或单眼一退一截术，随着时间的推移，患者远期手术成功率仍会显著下降，并出现不同程度的外斜视漂移现象^[8]。大型临床回顾性数据也表明，在初次水平眼外肌手术后，仍有 7%~8% 的患者因矫正不良面临二次手术风险^[9]，亟须引入更先进的辅助工具以突破现有诊疗瓶颈^[10]。

近年来，人工智能（特别是机器学习）在医学辅助决策中展现出巨大潜力^[11]，并在眼科图像识别及预后评估中取得了突破性进展^[12]。既往针对斜

视手术的机器学习研究主要集中于术后效果的回归预测。例如，de Almeida 等^[13]采用支持向量回归（support vector regression, SVR）预测水平斜视的单一眼肌干预量；Leite 等^[14]则探索了多输出回归树模型在斜视剂量预测中的应用。然而，这些系统存在明显的临床逻辑局限：其仅能在人为预先指定干预肌肉的前提下预测手术量，并未实现“肌肉选择”与“手术量计算”的联合决策。同时，诸如全连接神经网络等深度学习架构虽然在非线性映射上预测性能优异，但其“黑盒”特性缺乏可解释性，且将定性分类与定量回归混合处理，严重脱离了眼科医师“先决定干预哪条肌肉、再决定干预多少毫米”的真实诊疗思维^[15]。

为突破上述技术瓶颈，本研究旨在提出并构建一种贴合临床决策逻辑的递阶拓扑机器学习模型。本研究基于高维常规术前特征，将复杂的手术规划任务异构解耦为 2 个串联的阶段：首先通过分类节点智能推荐手术肌肉组合，随后通过回归节点条件输出精确的手术量。该方案期望在克服传统黑盒模型解释性差的基础上，实现较高精度的个体化手术规划，从而为规范斜视手术方案、降低不同资历医师间的决策变异性提供可靠的智能辅助依据。

1 资料与方法

1.1 一般资料

回顾性分析 2021 年 3 月—2024 年 3 月中南大学湘雅医院收治的 634 例行水平斜视手术患者的临床

资料。在进行数据质控时,针对临床边界进行严格过滤,剔除眼轴长度(axial length, AxL)、等效球镜及年龄等违背医学常理的 7 例极端异常录入数据后,最终纳入有效患者 608 例。患者平均年龄(18.38 ± 11.39)岁;平均眼轴长度(24.09 ± 1.46)mm。所有患者中右眼(oculus dexter, OD)为主斜眼者 259 例(42.6%),左眼(oculus sinister, OS)为主斜眼者 349 例(57.4%)。本研究共纳入外斜视患者 476 例,内斜视患者 132 例,分别作为外斜视组和内斜视组。纳入标准:①首次行水平直肌后徙或缩短手术;②年龄 2~65 岁。排除标准:①垂直性、麻痹性、限制性或旋转性斜视;②既往有斜视手术史;③手术记录单存在矛盾或核心字段缺失量>20%。本研究已获得中南大学湘雅医院医学伦理委员会批准(2026040693)。

1.2 仪器与设备

配有扫频源光相干断层扫描技术的 IOLMaster 700 型光学生物测量仪(德国 Zeiss 公司),Topcon KR-800 型全自动电脑验光仪(日本 Topcon 公司),采用 1~45 Δ 水平三棱镜(美国 Good-Lite 公司)、Sloan ETDRS 标准对数视力表进行检测,视力检测距离设定为 4 m,视力表符合 ISO 8596 国际标准要求。

1.3 递阶拓扑预测模型的构建

该研究采用 CatBoost 集成学习算法构建递阶拓扑模型,旨在模拟临床医师“先定性、后定量”的决策过程。

1.3.1 特征工程与变量衍生 从电子病历系统中提取 14 项常规术前指标。其中,利用双眼眼轴长度衍生计算平均眼轴长度及双眼眼轴差值;利用

双眼等效球镜衍生计算平均等效球镜及双眼球镜差值。

1.3.2 拓扑分级决策逻辑实现 模型拓扑结构分为 2 个层级。以分类模块作为第一阶结构,搭建含 8 个标签的多输出模型,用于预测双眼 4 条水平直肌各自对应的 8 类手术操作(后徙、缩短)是否需要实施手术干预;第二阶设置回归模块,仅对一阶判定为“干预”的样本开展手术矫正量的条件概率预测。模型初始输出结果经由临床互斥规则做后置校正处理,保障单眼内同一条肌肉不会同时执行后徙与缩短两类术式。

1.4 统计学方法

采用 Python 3.12 及 CatBoost 库进行模型训练与评估,硬件环境为 NVIDIA A100 计算集群。数据集按照 80% 训练集与 20% 测试集进行随机划分。分类指标采用宏平均曲线下面积(macro-average area under the curve, macro-AUC);回归指标采用平均绝对误差(mean absolute error, MAE)及均方根误差(root mean square error, RMSE)。计量资料以均数 $\pm$ 标准差($\bar{x} \pm s$)表示;计数资料以构成比(%)表示;相关性分析用 Pearson 法。 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 外斜视组与内斜视组临床资料比较

外斜视组与内斜视组年龄、平均眼轴、双眼平均等效球镜比较,经 t 检验,差异均无统计学意义($P > 0.05$)。外斜视组与内斜视组三棱镜度数比较,经 t 检验,差异有统计学意义($P < 0.05$)。见表 1。

表 1 外斜视组与内斜视组临床资料比较 ($\bar{x} \pm s$)

组别	<i>n</i>	年龄/岁	三棱镜度数/Δ	平均眼轴/mm	双眼平均等效球镜/D
外斜视组	476	18.25 ± 11.20	-54.74 ± 12.45	24.12 ± 1.50	-1.35 ± 2.65
内斜视组	132	18.85 ± 12.05	46.82 ± 11.35	23.98 ± 1.35	-1.10 ± 2.88
<i>t</i> 值		-0.536	-84.483	0.969	-0.941
<i>P</i> 值		0.592	0.000	0.333	0.347

2.2 术前三棱镜度数与主要眼外肌手术量的相关性分析

在单条肌肉手术频率统计中,左眼外直肌后徙术(lateral rectus recession of oculus sinister,

LRRcOS)执行频次最高(53.6%),其次为右眼外直肌后徙术(lateral rectus recession of oculus dexter, LRRcOD)(45.9%)。Pearson 相关性分析显示,术前三棱镜度数与各主要干预肌肉 LRRcOS、

LRRcOD、左眼内直肌缩短术 (medial rectus resection of oculus sinister, MRRsOS) 和右眼内直肌缩短术 (medial rectus resection of oculus dexter, MRRsOD) 的实际手术量均呈正相关 ($r=0.670$ 、 0.702 、 0.666 、 0.490 ，均 $P=0.000$)。见表 2。

表 2 单条肌肉手术干预频率与实际手术量分布

手术肌肉	例(%)	平均手术量/mm	手术量范围/mm
LRRcOS	326(53.6)	7.45	5.0 ~ 9.0
LRRcOD	279(45.9)	7.47	4.5 ~ 9.0
MRRsOS	237(39.0)	4.74	3.5 ~ 10.0
MRRsOD	181(29.8)	4.63	3.5 ~ 7.0

2.3 递阶拓扑模型预测性能评估

在 122 例测试集评估中，该递阶拓扑模型展现出优异的决策辅助性能。在拓扑第一阶（分类节点）任务中，模型对 4 块目标肌肉手术选择的 macro-AUC 达到 0.962，充分证明了其分类的准确性（见图 1）。其中左眼外直肌缩短术 (lateral rectus resection of oculus sinister, LRRsOS)、右眼外直肌缩短术 (lateral rectus resection of oculus dexter, LRRsOD)、右眼内直肌后徙术 (medial rectus recession of oculus dexter, MRRcOD)、左眼内直肌后徙术 (medial rectus recession of oculus sinister, MRRcOS)、LRRcOS、LRRcOD、MRRsOS 和 MRRsOD 单独分类的曲线下面积介于 0.908 ~ 0.995。在拓扑第二阶（回归节点）任务中，级联输出的手术量预测精度极高，MAE 为 0.491 mm，RMSE 为 1.494 mm。综合分类与回归预测结果，该模型生成的完整手术方案与专家执行方案的完全命中率达到 64.8% (79/122)。模型预测手术量与实际手术量的残差分布紧密围绕零线（见图 2）。

进一步的亚组分析显示，模型对外斜视与内斜视患者的完全命中率分别为 68.8% (66/96) 和 50.0% (13/26)，在儿童 (<18 岁) 与成人 (≥ 18 岁) 组中的完全命中率分别为 69.2% (45/65) 和 59.6% (34/57)。数据表明，模型在处理样本量占绝对优势的外斜视以及具有较高标准化干预特征的儿童病例时，表现出了极佳的决策鲁棒性；而对于样本量较小的内斜视群体及病程相对复杂的成人患者，其预测精度出现一定程度的衰减，这客观反映了机器学习模型对长尾数据分布的敏感

性，亦提示未来需在多中心研究中进一步扩充非典型病例的训练权重。

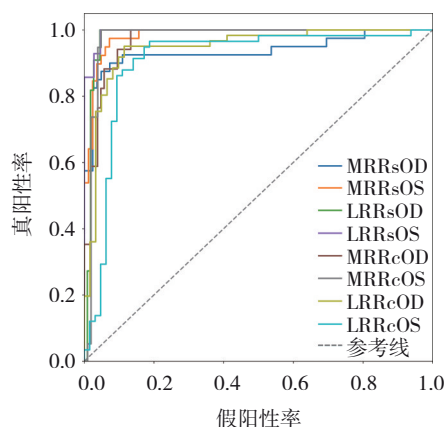


图 1 第一阶分类模型对各目标肌肉预测的受试者工作特征曲线

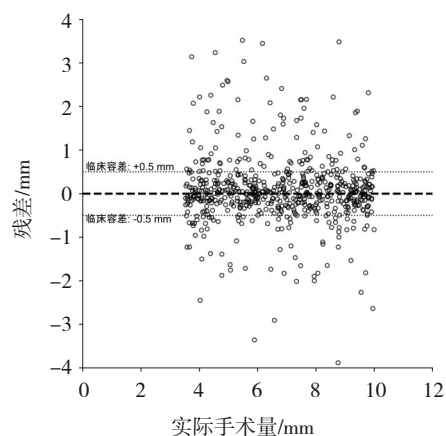


图 2 模型预测手术量与专家实际执行手术量的残差散点图

2.4 特征重要性分析

通过沙普利加和解释 (shapley additive explanations, SHAP) 法计算模型在训练过程中的贡献度得分，对 14 项术前临床特征的重要性进行了全局解释。结果显示，等效球镜指标是驱动模型进行手术规划的最核心特征，其权重显著高于其他指标；其次为术前三棱镜度数与主斜眼。这表明模型在决策过程中，不仅高度依赖客观的斜视度数，还深刻整合了双眼视力平衡状态及优势眼等高级临床因素。此外，单眼眼轴长度及双眼眼轴差值也表现出较高的权重，反映了解剖学差异及发育因素对最终定制化手术量的精细影响。

3 讨论

水平斜视作为临床最常见的斜视类型，其手

术方案及手术量的精准规划目前仍高度依赖专科医生的临床经验与主观判断。为了减少经验差异带来的不确定性,本研究突破了传统单一预测模型的局限^[16],创新性地提出并验证了一种基于 CatBoost 算法^[17]的两阶段递阶拓扑模型。

本研究的核心发现表明,该递阶拓扑模型在预测水平斜视的手术方案时展现出极高的精准度。在肌肉选择阶段,模型的 macro-AUC 达 0.962;在最终手术量预测阶段,其 MAE 仅 0.491 mm。在眼肌外科临床实践中,手术设计量通常以 0.5 mm 为最小调整量。本模型的预测误差低于 0.5 mm,意味着其计算出的手术量在绝大多数情况下可以直接转化为临床可执行方案。此外,完整方案的完全命中率高达 64.8%,表明该模型能够应对包含双眼多条肌肉联合干预的复杂手术设计。

回顾近年来该领域的研究进展,机器学习在斜视辅助决策中的应用大多停留在单一的回归预测层面^[18]。例如,de Almeida 等^[13]应用 SVR 预测水平斜视手术量,其报告的 MAE 在 0.5 ~ 0.8 mm;Leite 等^[14]使用多输出回归树模型,虽在多目标预测上有所尝试,但仍未能解决前置的“手术肌肉选择”问题。本研究不仅将手术量预测的 MAE 进一步降低至 0.491 mm,突破了临床可感知误差的界限,更在方法学与临床逻辑的契合度上具有显著优势。以往的研究往往将其视为“黑盒”映射,未能有效解耦“定性选择肌肉”与“定量手术量”的过程。本研究通过递阶拓扑设计,较好地模拟了临床医师“先定性、后定量”的真实诊疗思维。同时,结合 SHAP 可解释性分析^[15],模型实现了从“黑盒”向“白盒”的转变^[19],明确揭示了等效球镜^[20]、术前三棱镜度数及主斜眼在决策链条中的主导地位,极大地增强了临床应用的信任度。

本模型的成功建立具有重要的临床转化价值^[21]。由于斜视手术高度依赖术者的个人经验,不同资历或不同医疗机构的医师在面对同一患者时,可能给出存在明显差异的方案。引入该智能模型作为“第二意见”,有望规范区域内的手术规划标准,降低人为决策的变异性^[22],这与《中国儿童斜视围手术期管理专家共识(2025年)》^[4]中关于加强规范化与个体化术前评估的倡导高度契合,同时也可作为年轻眼科医师和专科规范化培训的优

质辅助示教工具^[23]。

然而,本研究仍存在一定局限性。首先,该研究数据源自单一中心的回顾性队列,可能包含特定专家的决策偏好,未来亟需开展多中心、前瞻性的外部验证以评估模型的泛化能力^[24]。其次,对于部分非典型的临床特征或发生率极低的极端手术量组合,受限于训练集样本量带来的长尾效应,模型的预测鲁棒性仍有进一步优化的空间。未来的研究将考虑引入更多维度的术前眼球运动轨迹视频及影像学参数^[25],以期进一步拓展人工智能模型在复杂性斜视(如伴有垂直或旋转偏斜)智能化规划中的应用边界。

综上所述,递阶拓扑机器学习模型能够高精度模拟临床专家的决策思维,实现较高精度的个体化手术规划,具有较强的临床决策支持潜力。

参考文献:

- [1] Hashemi H, Pakzad R, Heydarian S, et al. Global and regional prevalence of strabismus: a comprehensive systematic review and meta-analysis[J]. *Strabismus*, 2019, 27(2): 54-65.
- [2] 吴艳,冯雪亮,常敏,等.三棱镜递减疗法对小度数急性共同性内斜视的短期疗效观察[J]. *中华眼科杂志*, 2022, 58(3): 187-193.
- [3] Dagi L R, Velez F G, Archer S M, et al. Adult strabismus preferred practice pattern[®][J]. *Ophthalmology*, 2020, 127(1): P182-P298.
- [4] 中华医学会眼科学分会斜视与小儿眼科学组,中国医师协会眼科医师分会斜视与小儿眼科学组.中国儿童斜视围手术期管理专家共识(2025年)[J]. *中华眼科杂志*, 2025, 61(7): 500-506.
- [5] 许鑫悦,薛黎萍,滕凤舞.斜视术后立体视功能修复及重建的影响因素[J]. *中国现代医学杂志*, 2023, 33(18): 43-47.
- [6] Paduca A, Arnaut O, Bendelic E, et al. Extraocular muscle resection, recession length and surgery outcome modelling in strabismus treatment: a pilot study[J]. *BMJ Open Ophthalmol*, 2021, 6(1): e000802.
- [7] Kekunnaya R, Chandrasekharan A, Sachdeva V. Management of strabismus and its complexities[J]. *Asia Pac J Ophthalmol (Phila)*, 2023, 12(2): 185-195.
- [8] 韩秀敏.不同术式治疗儿童基本型或假性外展过强型间歇性外斜视的临床对照研究[J]. *中国现代医学杂志*, 2020, 30(11): 103-107.
- [9] Christensen M R, Pierson K, Leffler C T. Reoperation following strabismus surgery among Medicare beneficiaries: associations with geographic region, academic affiliation, surgeon volume, and adjustable suture technique[J]. *Indian J Ophthalmol*, 2018, 66(8): 1149-1153.
- [10] Moye A R, Jin L, Lambert S R. Machine learning in pediatric ophthalmology and strabismus[J]. *Curr Opin Ophthalmol*, 2023,

- 34(5): 417-422.
- [11] Topol E J. High-performance medicine: the convergence of human and artificial intelligence[J]. Nat Med, 2019, 25(1): 44-56.
- [12] Ting D S W, Peng L, Varadarajan A V, et al. Deep learning in ophthalmology: the technical and clinical considerations[J]. Prog Retin Eye Res, 2019, 72: 100759.
- [13] de Almeida J D S, Silva A C, Teixeira J A M, et al. Surgical planning for horizontal strabismus using Support Vector Regression[J]. Comput Biol Med, 2015, 63: 178-186.
- [14] Leite F H F, de Almeida J D S, da Cruz L B, et al. Surgical planning of horizontal strabismus using multiple output regression tree[J]. Comput Biol Med, 2021, 134(C): 104493.
- [15] Lundberg S M, Erion G, Chen H, et al. from local explanations to global understanding with explainable AI for trees[J]. Nat Mach Intell, 2020, 2(1): 56-67.
- [16] Wu C, Huang H, Zhao J, et al. A machine learning-based approach to predict the surgical outcome of intermittent exotropia[J]. BMC Ophthalmol, 2022, 22(1): 159.
- [17] Hancock J T, Khoshgoftaar T M. CatBoost for big data: an interdisciplinary review[J]. J Big Data, 2020, 7(1): 94.
- [18] Li J, Zhang Q, Liu J, et al. Artificial intelligence in strabismus: current and future applications[J]. Eye, 2023, 37(12): 2420-2426.
- [19] Amann J, Blasimme A, Vayena E, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective[J]. BMC Med Inform Decis Mak, 2020, 20(1): 310.
- [20] Buckley E G. The relationship between refractive error and strabismus[J]. Am Orthopt J, 2008, 58: 3-11.
- [21] Lu M, Peterseim M M W, Wilson M E, et al. Artificial intelligence-based approach to the screening of pediatric strabismus[J]. Am J Ophthalmol, 2021, 225: 158-164.
- [22] Zhang J, Wang Z, Zhao J, et al. Predicting the postoperative alignment of intermittent exotropia using machine learning[J]. Front Cell Dev Biol, 2021, 9: 778044.
- [23] Kazi R, Abolfathi A. Artificial intelligence in pediatric ophthalmology[J]. Curr Opin Ophthalmol, 2022, 33(5): 405-410.
- [24] Kelly C J, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence[J]. BMC Med, 2019, 17(1): 195.
- [25] Chen Z, Chen S, Zhang L, et al. A novel automated objective diagnostic system for strabismus utilizing deep learning and eye-tracking[J]. Front Cell Dev Biol, 2021, 9: 743845.
- (李科 编辑)
- 本文引用格式:** 王洁月, 吴小影. 基于递阶拓扑模型的斜视手术规划研究[J]. 中国现代医学杂志, 2026, 36(16): 63-68.
- Cite this article as:** Wang J Y, Wu X Y. Research on surgical planning for strabismus based on a hierarchical topological model[J]. China Journal of Modern Medicine, 2026, 36(16): 63-68.