

基于机器学习的人工耳蜗植入术后儿童听觉言语康复效果预测模型研究（附讲解视频）



扫码观看视频

白杰, 李颖, 金欣, 晏美棣, 刘海红

(国家儿童医学中心·首都医科大学附属北京儿童医院耳鼻咽喉头颈外科 北京 100045)

摘要 **目的:** 探讨机器学习技术在人工耳蜗(CI)植入术后儿童听觉言语康复效果预测中的应用。**方法:** 选取2012年1月—2024年10月首都医科大学附属北京儿童医院行CI植入术的187例儿童,收集其在开机时及开机后第1、3、6、12、24、36个月的父母评估儿童听说能力表现问卷数据及临床相关指标。运用机器学习算法(支持向量机、随机森林和人工神经网络)进行建模,并利用特征选择方法筛选影响听觉言语康复效果的重要影响因素。**结果:** 人工神经网络、随机森林和支持向量机三种机器学习方法构建预测模型的准确率分别为74.91%、71.02%、68.20%。经特征筛选,CI使用时间、开机月龄、性别、主要看护人受教育程度、居住地、干预方式、术前助听器使用史共7个特征具有显著性($P<0.05$)。**结论:** 机器学习技术可有效预测CI植入术后儿童听觉言语康复效果,为临床精准评估和个性化干预提供了新的工具与理论支持。

关键词 人工耳蜗; 机器学习; 听觉言语; 儿童

中图分类号 R764.5 **文献标识码** A **文章编号** 2096-7721(2025)04-0655-06

Prediction model based on machine learning for auditory and speech rehabilitation outcomes in children after cochlear implantation (with explanatory video)

BAI Jie, LI Ying, JIN Xin, YAN Meiling, LIU Haihong

(Department of Otolaryngology Head and Neck Surgery, Beijing Children's Hospital, Capital Medical University/National Center for Children's Health, Beijing 100045, China)

Abstract **Objective:** To explore the application of machine learning techniques in predicting auditory and speech rehabilitation outcomes for children after cochlear implantation. **Methods:** 187 children who underwent cochlear implantation at Beijing Children's Hospital Affiliated to Capital Medical University from January 2012 to October 2024 were selected. Data from the parents' evaluation of aural/oral performance of children questionnaire and clinical indicators were collected at device activation and 1, 3, 6, 12, 24, and 36 months after activation. Machine learning algorithms (Support Vector Machine, Random Forest, and Artificial Neural Network) were used to construct prediction models, with feature selection methods identifying key factors influencing rehabilitation outcomes. **Results:** The accuracy of prediction models constructed by Artificial Neural Network, Random Forest, and Support Vector Machine were 74.91%, 71.02%, and 68.20%, respectively. Feature selection revealed 7 significant predictors ($P<0.05$): usage time of CI, age at activation, gender, educational level of primary caregiver, residence location, cochlear implant laterality, and preoperative hearing aid use. **Conclusion:** Machine learning techniques can effectively predict auditory and speech rehabilitation outcomes in children after cochlear implantation, which provides a novel tool and theoretical support for precise clinical assessment and personalized intervention.

Key words Cochlear Implant; Machine Learning; Aural and Oral Performance; Children

世界卫生组织在《世界听力报告》^[1]中指出,全球超过15亿人存在一定程度的听力损失,其中至少4.3亿人需要专业的听力康复进行干预。对于听力障碍儿童而言,缺乏及时有效的干预将会对其言语发育、认知、社会心理健康、生活质量等产生负面影响^[2-4]。人工耳蜗(Cochlear Implant, CI)作为一种成熟的听力重建手段,已广泛应用于重度听力损失儿童^[5]。然而,CI植入术后的康复效果受年

基金项目: 国家重点研发计划项目(2023YFF1203504);北京市自然科学基金(7232059);高层次公共卫生技术人才建设专项(2022-3-016)
Foundation Item: National Key R & D Plan Project of China (2023YFF1203504); Natural Science Foundation of Beijing (7232059); High-level Public Health Technical Personnel Construction Project(2022-3-016)

引用格式: 白杰, 李颖, 金欣, 等. 基于机器学习的人工耳蜗植入术后儿童听觉言语康复效果预测模型研究(附讲解视频)[J]. 机器人外科学杂志(中英文), 2025, 6(4): 655-659, 666.

Citation: BAI J, LI Y, JIN X, et al. Prediction model based on machine learning for auditory and speech rehabilitation outcomes in children after cochlear implantation (with explanatory video) [J]. Chinese Journal of Robotic Surgery, 2025, 6(4): 655-659, 666.

通讯作者(Corresponding Author): 刘海红(LIU Haihong), Email: haihongliu6@aliyun.com

龄、家庭环境、听力损失程度等多种因素的影响^[6]。既往 CI 植入术后听觉言语康复效果研究主要使用统计学模型进行,难以全面探讨影响因素与结果之间的复杂关系,影响个性化干预方案的实施。近年来, AI 及其子领域机器学习算法在临床广泛应用,尤其在数据预测、特征筛选与个性化医疗方面^[7-9]。鉴于低龄儿童言语能力有限,父母评估儿童听说能力表现 (Parents' Evaluation of Aural/Oral Performance of Children, PEACH) 问卷已广泛用于患儿听觉语言能力评估^[10-11]。本研究应用机器学习算法并结合 PEACH 问卷数据,建立 CI 植入术后儿童听觉言语康复效果的预测模型,并筛选其关键影响因素,旨在为个性化临床干预提供科学支持。

1 资料与方法

1.1 一般资料 选取 2012 年 1 月—2024 年 10 月首都医科大学附属北京儿童医院行 CI 植入术的 187 例儿童。纳入标准:①符合《人工耳蜗植入工作指南 2013》中的植入标准^[12];②在开机时及开机后 1、3、6、12、24、36 个月完成术后随访及 PEACH 问卷评估。排除标准:存在神经发育异常、严重内耳畸形及其他可能严重影响听觉言语康复效果的疾病。

1.2 方法

1.2.1 资料收集 收集患儿相关临床指标及 PEACH 问卷结果,然后利用机器学习算法对问卷调查结果进行建模分析。PEACH 问卷^[11]是标准化量表,用于评估现实世界环境中在安静和嘈杂情况下,听力损失或听力正常儿童的功能性听觉能力和助听效果。在术后 1、3、6、12、24 和 36 个月对所有患儿进行了 PEACH 问卷评估,并收集问卷结果。临床相关指标包括:CI 使用时间、开机月龄、性别、主要看护人教育程度、居住地、术前听性脑干反应 (Auditory Brainstem Response, ABR) 阈值、术前听性稳态反应 (Auditory Steady-state Response, ASSR) 阈值、居住地及术前助听器佩戴史等。

1.2.2 数据处理

1.2.2.1 数据预处理与特征筛选:采用机器学习算法能够有效处理缺失值和剔除异常数据。在实际情况下,原始数据具有量纲差异,其数值范围不一。因此本研究对所有连续性变量进行了标准化处理。采用 Z-score 标准化方法,通过计算每个特征的均值 (μ) 和标准差 (σ),将原始数据转换为均值为 0、标准差为 1 的标准正态分布。计算公式为 $X' = (X - \mu) / \sigma$ 。其中, X 为原始数值, X' 为标准化后的数值。为了弥补不同分级出现分布不均衡的情况,本研究采用过

采样技术,对少数类样本进行人工少数类过采样法 (Synthetic Minority Oversampling Technique, SMOTE) 处理。特征筛选是机器学习中的关键步骤,旨在从原始特征集中识别并保留最具预测价值的变量,同时剔除冗余或无关特征。本研究纳入的数据特征包括干预方式、性别、开机月龄、术前 ABR 阈值、术前 ASSR 阈值、内耳畸形情况、术前助听器佩戴史、主要看护人受教育程度、居住地、CI 使用时间。利用 python 中 SelectKBest 函数对其进行特征选取,最终保留了 10 个最重要的特征,并呈现其 F 值图。

1.2.2.2 预测建模与验证:由于 PEACH 问卷第 1、2 个问题为设备日常使用和听觉舒适度问题,故不计入总分,其余 PEACH 问卷题目总分转化为百分比 (%),最后将 PEACH 问卷评估等级结果作为预测模型的目标变量。本研究采用人工神经网络 (Artificial Neural Network, ANN)、随机森林 (Random Forest, RF) 和支持向量机 (Support Vector Machine, SVM) 3 种经典机器学习算法构建预测模型。ANN 模型采用多层感知器结构,包含输入层、2 个隐藏层和输出层,采用 ReLU 激活函数和 Softmax 输出层;RF 模型采用集成多个决策树的预测结果;SVM 模型则采用径向基函数 (Radial Basis Function, RBF) 作为核函数。模型验证方法使用了五折交叉验证,训练集与测试集按 8 : 2 划分。本研究的 933 个数据经采样技术和数据集划分后,最终测试集数量为 283,呈现在混淆阵中。以上数据预处理、特征筛选、机器学习训练、模型验证使用 Python 3.0 Jupyter Lab 进行。

1.3 统计学方法 所有数据使用 SPSS 26.0 软件进行统计学分析,计数资料用例数 (百分比) [$n(\%)$] 表示,计量资料以均数 \pm 标准差 ($\bar{x} \pm s$) 表示,关键特征的显著性采用独立样本 t 检验, $P < 0.05$ 表示差异具有统计学意义。预测模型性能评价采用准确率、灵敏度、特异度及混淆矩阵。

2 结果

2.1 一般资料 本研究共纳入 187 例行 CI 植入术的儿童,在其术后 1、3、6、12、24 和 36 个月共收集 PEACH 问卷 933 份。PEACH 问卷评估结果按百分比分为 0%~25%、26%~50%、51%~75% 和 76%~100% 四个等级,结果显示第一级占 37.83%,第二级占 23.79%,第三级占 15.97%,第四级占 22.40%。患儿相关临床资料见表 1。

2.2 预测模型性能评价

2.2.1 预测模型性能评价指标 准确率、灵敏度和特异度是机器学习模型性能评价中的主要评价指标。

表 1 患儿一般资料 [$\bar{x} \pm s, n(\%)$]
Table 1 General information of children [$\bar{x} \pm s, n(\%)$]

指标	数值
开机月龄 (月)	22.81 ± 13.93
CI 使用时间 (月)	8.67 ± 12.22
性别	
男	107 (57.22)
女	80 (42.78)
内耳畸形情况	
正常发育	130 (69.52)
大前庭导水管综合征	57 (30.48)
干预方式	
单侧植入 CI	113 (60.43)
双侧植入 CI	55 (29.41)
双模式 (一侧 CI, 一侧助听器)	19 (10.16)
ABR 阈值 (dB nHL)	89.26 ± 9.32
ASSR 阈值 (dB nHL)	91.39 ± 13.91
术前助听器佩戴史	
有	27 (14.44)
无	160 (85.56)
主要看护人受教育程度	
小学	73 (39.03)
初中	47 (25.13)
高中	31 (16.58)
大专	18 (9.63)
本科	15 (8.02)
研究生	3 (1.60)
居住地	
乡村	110 (58.82)
小城市	33 (17.65)
中等城市及大城市	44 (23.53)

其中, 准确率作为总体性能度量指标, 其表达式为: $(TP+TN) / (TP+TN+FP+FN)$ 。真阳性 (True Positive, TP) 表示正确识别第 i 级样本的数量; 真阴性 (True Negative, TN) 为正确识别非 i 级样本的数量; 假阳性 (False Positive, FP) 为错误地将非 i 级样本判为 i 级的数量; 假阴性 (False Negative, FN) 为错误地将 i 级样本判为其他级别的样本数量。灵敏度计算公式为: $TP / (TP+FN)$, 衡量模型识别目标级别的能力; 特异度计算公式为: $TN / (TN+FP)$, 衡量模型排除非目标级别的能力。

本研究中 ANN 模型整体准确率最高, 为 74.91%, 表现出较强的分类能力。该模型在 PEACH 问卷各个分级的灵敏度、特异度均较为均衡, 尤其是在第三级的灵敏度 (78.26%) 和特异度 (90.19%) 较高, 表现出最佳的预测效果。而 RF 模型整体准确率为 71.02%, 在 PEACH 问卷第四级表现最佳, 具有较高的灵敏度 (79.45%), 且在该分级特异度最高 (93.81%)。然而, 该模型在 PEACH 问卷第二级的灵敏度较低, 仅为 52.78%。SVM 模型整体准确率相对较低 (68.20%)。该模型在 PEACH 问卷第一级和第四级的灵敏度相对较高, 其第四级灵敏度和特异度达到 76.71% 和 92.38%。综合而言, 三种模型在 PEACH 问卷第二级中的灵敏度均弱于其他级, 在第四级的特异度优于其他级 (见表 2)。

2.2.2 预测模型混淆矩阵 混淆矩阵是评估分类模型性能的重要工具, 直观展示预测结果与真实结果之间的对应关系。其中, 对角线元素表示正确分类的样本数量, 而非对角线元素则表示被错误分类的样本数量。

本研究 ANN 算法所构建的模型表现出较好的分类能力。混淆矩阵 (如图 1A) 数据计算结果显示, 第一级、第三级和第四级的分类准确率分别为 53/69 (76.81%)、54/69 (78.26%) 和 57/73 (78.08%), 而第二级的准确率为 48/72 (66.67%)。第二级样本被误分为第一级的情况较为显著, 第一级与第二级之间存在互误现象。

表 2 三种预测模型性能评价
Table 2 Performance evaluation of three prediction models

模型	第一级 (0%~25%)		第二级 (26%~50%)		第三级 (51%~75%)		第四级 (76%~100%)		整体准确率 (%)
	灵敏度	特异度	灵敏度	特异度	灵敏度	特异度	灵敏度	特异度	
人工神经网络模型	76.81	91.12	66.67	90.05	78.26	90.19	78.08	95.24	74.91
随机森林模型	79.91	89.72	52.78	90.52	72.46	87.38	79.45	93.81	71.02
支持向量机模型	75.36	91.12	63.24	85.58	57.53	88.57	76.71	92.38	68.20

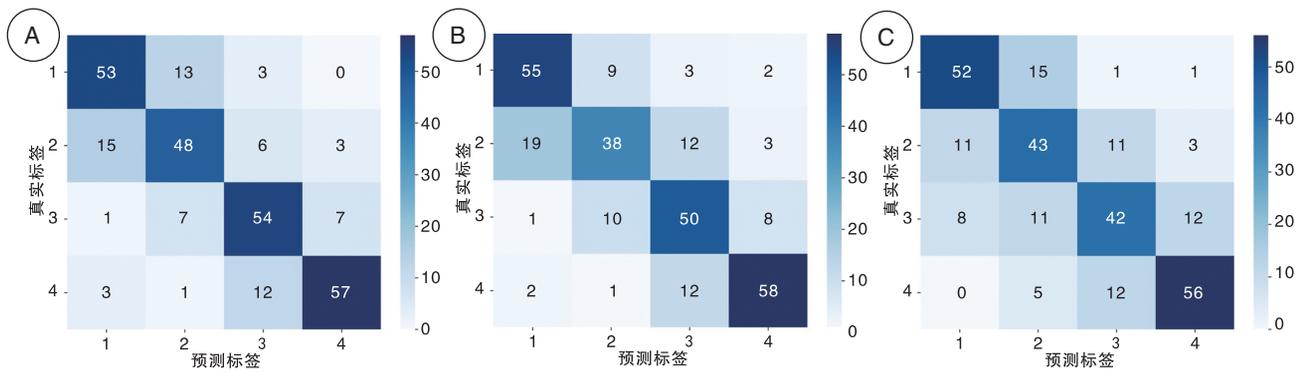


图 1 混淆矩阵

Figure 1 Confusion matrix

注: A. 神经网络模型; B. 随机森林模型; C. 支持向量机模型

RF 模型呈现出良好的分类表现。混淆矩阵 (如图 1B) 数据计算结果显示, 第一级、第三级和第四级的分类准确率分别为 55/69 (79.71%)、50/69 (72.46%) 和 58/73 (79.45%), 表现较为优异; 而第二级的准确率为 38/72 (52.78%), 相对偏低。第二级样本存在较高的误分率, 尤其是被误分为第一级的情况较多 (19 例), 这表明模型在区分第一级与第二级样本时存在一定挑战。第三级样本被误分为第四级 (8 例) 和第四级样本被误分为第三级 (12 例) 的情况也较为突出, 这显示这两个级别之间的边界特征需要进一步优化。

SVM 模型展现出较为稳定的分类表现。混淆矩阵 (如图 1C) 数据计算结果证实了其良好的分类能力。第一级、第二级、第三级和第四级的分类准确率分别为 52/69 (75.36%)、43/68 (63.24%)、42/73 (57.53%) 和 56/73 (76.71%)。值得注意的是, 第三级样本的分类准确率相对较低, 主要原因在于有 12 例第三级样本被误分为第四级, 同时也有 12 例第四级样本被误分为第三级, 表明模型在区分这两个级别时存在一定困难。此外, 第二级与第三级之间也存在互误现象, 各有 11 例样本被错误分类。

2.3 重要特征 本研究中纳入了 CI 使用时间、干预方式、性别、开机月龄、术前 ABR 阈值, 术前 ASSR 阈值、内耳畸形情况 (正常发育和前庭导水管扩大)、术前助听器佩戴史、主要看护人受教育程度、居住地共 10 个变量作为模型特征。其中, 将开机月龄作为分类变量纳入模型, 将研究对象分为三组: 早期植入组 (<12 月龄)、中期植入组 (12~24 月龄) 及晚期植入组 (>24 月龄)。经过特征筛选后具有显著性 ($P < 0.05$) 的重要特征有 7 个, 分别为 CI 使用时间、开机月龄分组、性别、主要看护人受教育程度

程度、居住地、干预方式、术前助听器使用史。其中 CI 使用时间、开机月龄分组的 F 值极高 ($P < 0.001$), 这显示其对模型具有极大的影响。特征的 F 值与 P 值如图 2。

3 讨论

本文基于神经网络、随机森林和支持向量机三种机器学习方法构建了预测模型, 准确率分别达到 74.91%、71.02%、68.20%。特征筛选过程发现, 有 7 个因素显著影响 CI 植入术后儿童听觉语言康复效果, 包括 CI 使用时间、开机月龄、性别、主要看护人教育程度、居住地、干预方式、术前助听器使用史。本研究使用 PEACH 问卷和机器学习算法预测 CI 植入术后儿童听觉语言康复效果, 为 CI 植入术的个性化调试提供了快速无创的新工具和新方法。

目前, 国内外已初步探索了 CI 植入效果预测模型, 然而其效果尚待进一步的外部验证。有研究表明这些预测模型的结局指标存在显著异质性, 包括言语识别率、标准化问卷评分以及研究者自行

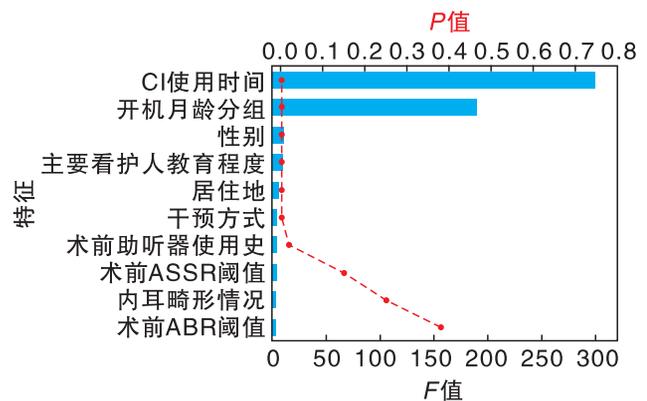


图 2 重要特征 (按 F 值排序)

Figure 2 Important features arranged by F -value from high to low

构建的综合指标等，其分类预测模型的准确度区间为70%~95%，与本研究所获得的准确率水平相当^[13-14]。Carlson M L等人^[15]开发的机器学习模型可依据术前数据预测言语识别得分，准确率可达87%，敏感度和特异度达90%和80%。LU S等人^[16]根据70例CI植入患者的术前检查及术后2年的随访数据构建了预测模型，其对听觉行为分级标准（Categories of Auditory Performance, CAP）和言语可懂度分级（Speech Intelligibility Rating, SIR）得分的预测准确率达71%和93%。一项以言语识别评分为主要结局指标的预测模型共纳入2489例CI植入者^[13]，结果表明机器学习算法较传统统计模型显著提升了预测效能（ $P<0.001$ ）。多项研究通过整合神经解剖学特征或影像学数据，进一步优化了模型的预测准确性与临床适用性。Feng G等人^[17]的研究加入神经形态学数据作为特征，模型准确率达到76%。Tan L等人^[18]则利用CI植入前功能性核磁共振数据预测CI植入两年后患儿的言语技能发展情况，二分类效果达到了81.3%。WENG J等人^[19]利用耳蜗解剖学参数构建内耳畸形CI植入儿童预测模型，其对CAP问卷和SIR问卷得分的准确率可达93.3%~86.7%。本研究未来有望引入功能性核磁共振、脑电图等神经影像学数据，以进一步提高预测准确度。

影响儿童CI植入效果的因素众多，植入年龄、特殊疾病（如听神经病、蜗神经发育不良）、康复训练情况、术前听觉语言能力等是研究关注的重点^[6, 20]。目前国内外研究^[4, 20]已经证实，植入年龄和CI使用时间对术后听觉语言能力评估有显著影响。多数研究显示，CI植入年龄越小，康复效果越好。早期植入被认为是促进患儿词汇表达、言语发展的关键因素^[21]。本研究中居住地（乡村、城镇、城市）和主要看护人受教育程度也是影响CI植入术后患儿听觉语言能力的显著因素，与Ching T Y等人的研究类似^[11, 20]。值得注意的是，本研究发现性别是显著的影响因素，而冀飞等人^[22]的研究则认为性别非显著影响因素，这可能是由于机器学习的数据驱动特性，其不预先对数据做出假设，但能从数据中挖掘出潜在规律。文中术前助听器使用史（14.44%）和双模式植入（4.81%）的患儿比例偏低，可能影响了特定特征在模型中的权重估计精确度，从而限制了特征筛选的全面性和稳健性。除在术后康复效果预测外，AI在CI领域方面有着广阔应用前景。AI已应用于CI编码策略中，提高了CI在复杂声学环境中的性能表现^[23-24]。CI作为日常使用的设备，可产

生大量的个体用户数据，持续收集用户行为、偏好等个体数据用于AI分析，可进一步优化CI使用者的日常体验。此外，AI还可应用于CI调机，有研究^[25]表明，与单独手动调试相比，利用AI作为辅助工具可改善听力效果。

综上所述，CI使用时间、开机月龄、性别、主要看护人受教育程度、居住地等对术后听觉语言能力评估有显著影响。然而，本研究尚未开展严格的外部验证，模型在不同人群和医疗环境中的泛化能力有待进一步评估。未来研究应扩大样本量，并通过多中心外部验证强化模型的稳健性和可信度，以期为临床决策提供更为可靠的辅助工具。

利益冲突声明：本文不存在任何利益冲突。

作者贡献声明：白杰负责数据分析，绘制图表，设计论文框架，起草论文；白杰、李颖、金欣、晏美根、刘海红均参与该项目具体操作及研究过程的实施；李颖、金欣、晏美根负责数据收集与整理；刘海红负责拟定写作思路，指导撰写文章并最后定稿。

参考文献

- [1] World Health Organization. World report on hearing[R/OL]. Geneva: WHO, 2021[2021-04-30]. <https://www.who.int/publications/i/item/world-report-on-hearing>.
- [2] Olusanya B O, Neumann K J, Saunders J E. The global burden of disabling hearing impairment: a call to action[J]. Bull World Health Organ, 2014, 92(5): 367-373.
- [3] Nordvik Ø, Laugen Heggdal P O, Brännström J, et al. Generic quality of life in persons with hearing loss: a systematic literature review[J]. BMC Ear Nose Throat Disord, 2018. DOI: 10.1186/s12901-018-0051-6.
- [4] Shield B. Evaluation of the social and economic costs of hearing impairment[R/OL]. A report for Hear, 2006[2021-04-30]. https://www.hear-it.org/de/multimedia/Hear_It_Report_October_2006.pdf.
- [5] Carlyon R P, Goehring T. Cochlear implant research and development in the twenty-first century: a critical update [J]. Journal of the Association for Research in Otolaryngology: JARO, 2021, 22(5): 481-508.
- [6] Cosetti K, Waltzman S B. Outcomes in cochlear implantation: variables affecting performance in adults and children [J]. Otolaryngologic Clinics of North America, 2012, 45(1): 155-171.
- [7] Frosolini A, Franz I, Caragli V, et al. Artificial intelligence in audiology: a scoping review of current applications and future directions [J]. Sensors (Basel), 2024, 24(22): 7126.
- [8] Beam A L, Drazan J M, Kohane I S, et al. Artificial intelligence in medicine [J]. The New England Journal of Medicine, 2023, 388(13): 1220-1221.
- [9] 魏兴梅, 薛书锦, 高振橙, 等. 人工智能在人工耳蜗植入中的应用 [J]. 首都医科大学学报, 2024, 45(6): 931-937.
- [10] 郭倩倩, 陈雪清, 孟超, 等. 常用听觉能力评估问卷临床应用比较 [J]. 中国耳鼻咽喉头颈外科, 2018, 25(9): 485-489.
- [11] Ching T Y, Hill M. The Parents' Evaluation of Aural/Oral Performance of Children (PEACH) scale: normative data [J]. Journal of the American Academy of Audiology, 2007, 18(3): 220-235.
- [12] 中华耳鼻咽喉头颈外科杂志编辑委员会, 中华医学会耳鼻咽喉头颈外科学分会, 中国残疾人康复协会听力语言康复专业委员会. 人工耳蜗植入工作指南 (2013) [J]. 中华耳鼻咽喉头颈外科杂志, 2014, 49(2): 89-95.

（下转 666 页）