



DOI:10.11817/j.issn.1672-7347.2022.210704

基于3D卷积神经网络的中耳疾病高分辨率CT图像 辅助分类诊断模型的应用

粟日¹, 宋剑^{2,3,4}, 汪政¹, 卯爽^{2,3,4}, 毛弈韬⁵, 吴学文^{2,3,4}, 侯木舟¹

[1. 中南大学数学与统计学院, 长沙 410083; 2. 中南大学湘雅医院耳鼻咽喉科, 长沙 410008;
3. 耳鼻咽喉重大疾病湖南省重点实验室, 长沙 410008; 4. 国家老年疾病临床医学研究中心(湘雅医院),
长沙 410008; 5. 中南大学湘雅医院影像科, 长沙 410008]

[摘要] 目的: 慢性化脓性中耳炎(chronic suppurative otitis media, CSOM)和中耳胆脂瘤(middle ear cholesteatoma, MEC)是两类临床上最常见的慢性中耳疾病。在诊疗过程中, 该两类疾病因具有类似的临床表现, 容易造成误诊及漏诊。高分辨率计算机断层扫描(high resolution computed tomography, HRCT)能清晰地显示颞骨的精细解剖结构, 准确地反映中耳病变情况及病变范围, 对慢性中耳疾病的鉴别诊断具有优势。本研究开发一种基于颞骨HRCT影像数据, 对慢性中耳疾病实施自动信息提取与分类诊断的深度学习模型, 旨在提高临床上对慢性中耳疾病的分类诊断效率, 减少漏诊及误诊的发生。**方法:** 回顾性收集2018年1月至2020年10月于湘雅医院耳鼻咽喉科住院的慢性中耳疾病患者的临床病历及颞骨HRCT影像资料。由2名经验丰富的耳鼻咽喉科医师独立审查患者的医疗记录, 并对最终诊断达成一致结论。最终纳入499例患者(998侧耳), 将998侧耳分为3组: MEC组(108侧耳)、CSOM组(622侧耳)、正常组(268侧耳)。使用不同方差的高斯噪声进行数据集样本扩增处理, 以此消除组间样本数量的不平衡。经扩增后的实验数据集样本量为1 806侧耳, 实验中随机选择75%(1 355侧耳)用于训练, 10%(180侧耳)用于验证, 剩余的15%(271侧耳)用于测试并评估模型性能。模型整体设计为串联式结构, 设置具有3种不同功能的深度学习模型: 第一种是区域推荐网络算法, 从整体HRCT图像中搜索中耳部分的图像进行切割、保存; 第二种是基于孪生网络结构的图像对比卷积神经网络(convolutional neural network, CNN), 从切割好的图像中搜索与HRCT图像关键层面匹配的图像, 并进行3D数据块的构建与保留; 第三种是基于3D-CNN操作, 用于对3D数据块进行分类诊断, 并给出最后的预测概率。**结果:** 基于孪生网络结构的特殊层面搜索网络在10个特殊层面上表现出了0.939的平均AUC值。基于3D-CNN的分类网络整体准确率为96.5%, 整体召回率为96.4%, 3种判类结果的平均AUC值为0.983。预测结果中的CSOM病例召回率为93.7%, MEC病例召回率为97.4%。在后续进行的对比实验上, 一些经典的CNN平均精确度为79.3%, 平均召回率为87.6%。本研究构建的深度学习网络的准确度比普通的CNN提升约17.2%, 召回率提升约8.8%。**结论:** 本研究构建的深度学习网络模型可以自动从患者颞骨HRCT影像数据中提取含有中耳特征的3D数据块, 在降低数据整体规模的同时保存了对应图像间的关系, 可进一步使用3D-CNN进行CSOM与MEC的分类诊断。该模型的设计很好地结合了HRCT数据的连续性, 实验结果准确度高, 适应性广, 优于目前常用的CNN方法。

[关键词] 慢性化脓性中耳炎; 中耳胆脂瘤; 孪生网络; 3D卷积神经网络

收稿日期(Date of reception): 2022-04-11

第一作者(First author): 粟日, Email: suricsu@csu.edu.cn, ORCID: 0000-0003-2516-9243

通信作者(Corresponding author): 宋剑, Email: entsj@csu.edu.cn, ORCID: 0000-0002-8447-8647

基金项目(Foundation item): 国家自然科学基金(81700923); 中国博士后科学基金(2021M693566, 2021T140751); 湖南省自然科学基金(2021JJ31108, 2021JJ41017); 湖南省科技创新人才计划(2020RC2013)。This work was supported by the National Natural Science Foundation (81700923), the Postdoctoral Science Foundation (2021M693566, 2021T140751), the Natural Science Foundation of Hunan Province (2021JJ31108, 2021JJ41017), and the Science and Technology Innovation Program of Hunan Province (2020RC2013), China.

Application of high resolution computed tomography image assisted classification model of middle ear diseases based on 3D-convolutional neural network

SU Ri¹, SONG Jian^{2,3,4}, WANG Zheng¹, MAO Shuang^{2,3,4}, MAO Yitao⁵, WU Xuewen^{2,3,4}, HOU Muzhou¹

(1. School of Mathematics and Statistics, Central South University, Changsha 410083;

2. Department of Otorhinolaryngology, Xiangya Hospital, Central South University, Changsha 410008;

3. Hunan Provincial Key Laboratory of Major Otorhinolaryngology Diseases, Changsha 410008;

4. National Clinical Research Center for Geriatric Diseases, Xiangya Hospital, Changsha 410008;

5. Department of Imaging, Xiangya Hospital, Central South University, Changsha 410008, China)

ABSTRACT

Objective: Chronic suppurative otitis media (CSOM) and middle ear cholesteatoma (MEC) are the 2 most common chronic middle ear diseases. In the process of diagnosis and treatment, the 2 diseases are prone to misdiagnosis and missed diagnosis due to their similar clinical manifestations. High resolution computed tomography (HRCT) can clearly display the fine anatomical structure of the temporal bone, accurately reflect the middle ear lesions and the extent of the lesions, and has advantages in the differential diagnosis of chronic middle ear diseases. This study aims to develop a deep learning model for automatic information extraction and classification diagnosis of chronic middle ear diseases based on temporal bone HRCT image data to improve the classification and diagnosis efficiency of chronic middle ear diseases in clinical practice and reduce the occurrence of missed diagnosis and misdiagnosis.

Methods: The clinical records and temporal bone HRCT imaging data for patients with chronic middle ear diseases hospitalized in the Department of Otorhinolaryngology, Xiangya Hospital from January 2018 to October 2020 were retrospectively collected. The patient's medical records were independently reviewed by 2 experienced otorhinolaryngologist and the final diagnosis was reached a consensus. A total of 499 patients (998 ears) were enrolled in this study. The 998 ears were divided into 3 groups: an MEC group (108 ears), a CSOM group (622 ears), and a normal group (268 ears). The Gaussian noise with different variances was used to amplify the samples of the dataset to offset the imbalance in the number of samples between groups. The sample size of the amplified experimental dataset was 1 806 ears. In the study, 75% (1 355) samples were randomly selected for training, 10% (180) samples for validation, and the remaining 15% (271) samples for testing and evaluating the model performance. The overall design for the model was a serial structure, and the deep learning model with 3 different functions was set up. The first model was the regional recommendation network algorithm, which searched the middle ear image from the whole HRCT image, and then cut and saved the image. The second model was image contrast convolutional neural network (CNN) based on twin network structure, which searched the images matching the key layers of HRCT images from the cut images, and constructed 3D data blocks. The third model was based on 3D-CNN operation, which was used for the final classification and diagnosis of the 3D data block construction, and gave the final prediction probability.

Results: The special level search network based on twin network structure showed an average AUC of 0.939 on 10 special levels. The overall accuracy of the classification network based on 3D-CNN was 96.5%, the overall recall rate was 96.4%, and the average AUC under the 3 classifications was 0.983. The recall rates of CSOM cases and MEC cases were 93.7% and 97.4%, respectively. In the subsequent comparison experiments, the average accuracy of some classical CNN was 79.3%, and the average recall rate was 87.6%. The precision rate and the recall rate of the deep learning network constructed in this study were about 17.2% and 8.8% higher than those of the common CNN.

Conclusion: The deep learning network model proposed in this study can automatically extract 3D data blocks containing middle ear features from the HRCT image data of patients' temporal bone, which can reduce the overall size of the data while preserve the relationship between corresponding images, and further use 3D-CNN for classification and diagnosis of CSOM and MEC. The design of this model is well fitting to the continuous characteristics of HRCT data, and the experimental results show high precision and adaptability, which is better than the current common CNN methods.

KEY WORDS chronic suppurative otitis media; middle ear cholesteatoma; siamese network; 3D convolutional neural network

中耳位于颞骨岩部,由鼓膜、听骨链、鼓室、咽鼓管及乳突等结构组成。中耳的主要功能是将空气中的声波振动能量高效地传递到内耳淋巴液,其中鼓膜和听骨链在声音传递过程中起重要作用^[1]。由于其特殊的解剖结构,中耳疾病在耳部疾病中的占比最高,也是造成听力损失的主要原因之一^[2]。其中以慢性化脓性中耳炎(chronic suppurative otitis media, CSOM)及中耳胆脂瘤(middle ear cholesteatoma, MEC)两类疾病最常见。CSOM是一类产生于中耳黏膜、骨膜或深达骨质的慢性化脓性炎症^[3]。而MEC是一种位于中耳内的囊性结构,囊的内壁为复层鳞状上皮,产生角蛋白的鳞状上皮异常积聚在中鼓室、上鼓室、乳突、岩尖等部位^[4]。这两类疾病均可表现为反复耳溢液、不同程度的听力下降及耳鸣等,具有一定相似性。但二者的治疗目的与原则需要区分开来:MEC首选手术治疗,尽可能完整地清除病灶;而CSOM则可根据病变情况选择药物治疗或手术干预。故二者的鉴别诊断对于患者的临床诊疗和预后判断具有重要意义^[4-5]。

颞骨高分辨率计算机断层扫描(high resolution computed tomography, HRCT)是目前诊断中耳疾病的一种相对准确的影像学方法,其能清晰地显示中耳内部解剖结构和病变,加之多种图像后处理技术的开发和利用,能从三维(three dimensions, 3D)角度展现颞骨精细的解剖结构。由于颞骨HRCT检测中耳病变的灵敏度高,目前已成为中耳疾病影像检查的

首选方法^[6]。

CSOM和MEC的HRCT图像均可表现为中耳腔和乳突腔内的软组织肿块样低密度影部分填充。而MEC的典型HRCT影像可出现中耳听力结构和相邻颅骨的进行性破坏,如上鼓室盾板结构破坏、鼓室窦口扩大等^[7]。本研究基于CSOM和MEC的HRCT影像学特点,结合HRCT图像数据天然的连续性,利用计算机视觉算法优化图像网络数据输入规模,并采用3D卷积神经网络(3D convolutional neural network, 3D-CNN)构建一种人工智能辅助快速诊断模型,以期用该模型从患者颞骨HRCT扫描数据中自动筛选并构造出具有中耳结构特征的3D数据块,最终通过3D-CNN来完成诊断。

1 资料与方法

1.1 数据来源

本研究经中南大学湘雅医院医学伦理委员会批准(审批号:201703336)。收集2018年1月至2020年10月于湘雅医院耳鼻咽喉科住院实施中耳手术患者的病历资料及颞骨HRCT影像资料。由2名具有副主任医师职称的耳鼻咽喉科医师(年资均大于15年)独立审查所有患者的医疗记录并达成一致结论。经对病历进行筛查,排除先天性中耳畸形、再次手术患者及急慢性分泌性中耳炎的患者,最终纳入499例患者(998侧耳)进行本次研究。患者年龄5~72(38.75±

14.38)岁,其中男199例,女300例。颞骨HRCT扫描的水平位影像可以显示双耳,998侧耳被分为3组:MEC组(108侧耳)、CSOM组(622侧耳)、正常组(268侧耳)。所用数据集的内部信息见表1。

表1 数据集的构造

Table 1 Structure of dataset

组别	左耳/侧	右耳/侧	合计/侧
CSOM组	314	308	622
MEC组	46	62	108
正常组	139	129	268
合计	499	499	998

CSOM: 慢性化脓性中耳炎; MEC: 中耳胆脂瘤。

用256排 Revolution CT扫描仪(美国GE Healthcare公司)进行颞骨HRCT检查。扫描参数为:管电压100 kV,管电流325 mA,螺距系数0.688,矩阵512×512,视场20 cm,层厚0.625 mm。HRCT扫描体位为标准头颅前后位,扫描方式为螺旋扫描,窗宽3 000~4 000 HU,窗位为300~500 HU。对于已进行手术的患者,最终的诊断结论是依据术后病理

结果;而未手术的正常耳的诊断,则由专科医师结合病史、体检、听力学检查和影像学检查综合评估。在所使用的病例数据中,存在22侧耳是MEC合并慢性中耳炎,其优先被诊断为MEC。

1.2 模型运行硬件环境与设定

整体模型的所有实验均在个人服务器(i5-10400F六核2.90 GHz处理器、64 GB RAM和一个NVIDIA GeForce GTX 2070s显卡)上进行,在Python中使用PyTorch框架实现。实验中其他超参数设定为固定值,其中学习率设定为0.0001,训练周期epoch设定为100,训练批次大小设定为100。

1.3 图像预处理

在专科医师的指导下,对HRCT图像原始数据进行系统的手工标注。选取中耳结构作为目标值,用一个100像素×100像素的框裁剪原有HRCT图像,然后按照裁剪区域作为方案头部算法输入的有效图像分析区域。图1为手工标记的示例。对于分类模型,根据数据的后缀名生成与处理后图像相对应的CSV文件,以获得正确的诊断标签。

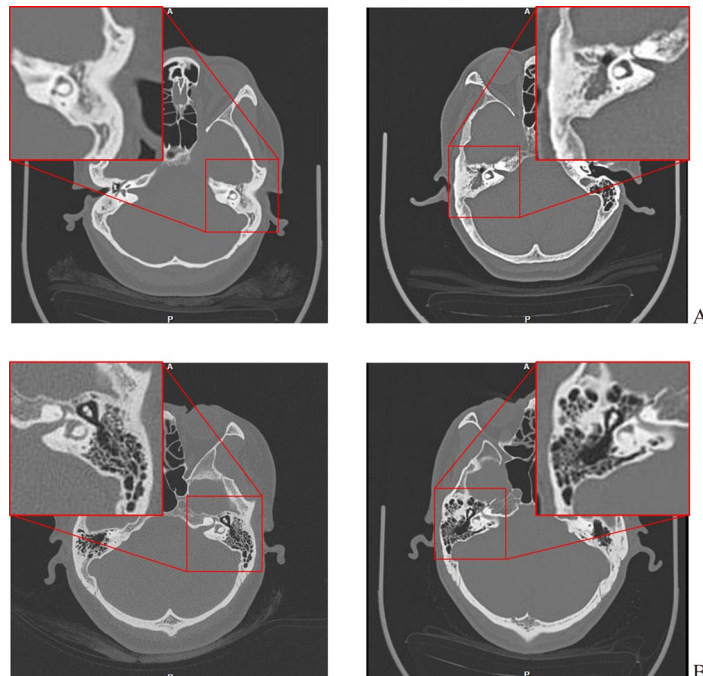


图1 单侧中耳结构数据标记示例

Figure 1 Example of data labeling for a middle ear structure

A: Schematic representation of middle ear markers in normal adults; B: Schematic representation of middle ear markers in patients with chronic suppurative otitis media.

由于3组病例的数据量不对等,在整体数据集的占比上存在严重偏差,导致数据集呈现一种严重的数据不平衡的现象,可能出现模型更偏向于学习大样本类别的数据,从而导致小样本类别的数据在测试或实际应用中无法得到充分的解析与诊断。为了应对数量偏差,本研究对原始图像数据集进行了高斯滤波处理^[8]。通过正态分布来调整每个像素点上的数值,以达到微量干扰图像质量的效果。在实现上,高斯噪声利用设定好均值和方差的正态分布取随机数,并对图像每个像素点进行归一化(将0~255的整数压缩至0~1的浮点数),加上随机数引入噪声干扰(将取到负值的像素点取值为0),然后将像素值放缩使之回至常规范围,最后保存处理后的图像。基于数据不平衡问题,本研究使用不同的方差来对原始图像数据集进行数量上的增强。除了使用高斯噪声,在实现中构建结构性数据集时还使用了随机细小角度倾斜图像、翻转图像及在图像边缘进行细微切割等常规数据增强手段,以此来增大算法所接受的数据量的整体规模。这样的数据增强方法可使模型在学习小样本类别的数据时具有一定的泛化性。

1.4 模型设计

本研究中诊断模型的基本思想是对原始HRCT图像数据进行自助辅助诊断。在分析数据时,笔者发现水平位高密度HRCT扫描图像具有层数多、存在上下连续性且图像噪声大等特性。因此本研究将模型整体设计分为3个部分:1)使用区域推荐网络对每个HRCT图像的中耳部分进行识别与切割,以此来去除不需要的图像以及图像中非中耳区域的信息;2)使用基于孪生网络结构的图像对比网络,从切割出来的图像中筛选出与特殊层面相似度高的目标图像并构建3D特征数据块,以此来解决数据规模过大的问题,在保证图像上下连续性的同时降低整体输入数据的规模;3)针对构建的3D图像数据块,基于3D-CNN的方法,利用HRCT数据的天然连续性,对最后的分类置信度进行输出,完成最后的诊断工作。图2为模型的整体结构,图3为模型中部分模块的细节设置。

1.5 区域推荐网络算法

本模型使用的自动提取特征图像方法为区域推荐网络(region proposal network, RPN)算法。这种算法是已广泛应用于计算机视觉目标检测任务中的区域卷积神经网络(region convolutional neural network, R-CNN)系列方法^[9-10]。该方法在输入图像上通过设置待推荐区域的锚点,通过卷积过程中在特征图上的

映射得到大量的备选区域,然后输出大多数区域的分类概率值(基于训练样本),并选择概率最大的区域作为感兴趣区域。整体RPN计算过程可以分为两大部分:第一部分为锚点的寻找;第二部分为候选框的设定。在构建此模块时,本研究沿用了R-CNN系统方法中所设定的损失函数:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

其中, i 为批量训练中的第 i 个锚点,表示第 i 个锚点为前景的概率, $p_i^* \in \{0,1\}$ 为锚点是否为前景的真实概率, t_i 表示预测的选区框的坐标, t_i^* 为选区框的真实标记。

对于RPN模块的设计,本研究根据原始HRCT图像的大小将选择框设置为100像素×100像素的固定大小,并确定中耳结构位于图像的中心,对本研究所构建的中耳特征图进行标准化。

为了提取颞骨HRCT扫描中耳特征图像的两侧,笔者对图像的输出进行了不同的概率判别,以此来完成对于图像左侧(右耳)和右侧(左耳)的同时输出。在此设置中,可以降低该部分的时间复杂度,同时可以通过调整选择特征图的概率阈值来确保非中耳结构的特征图像不会选择到区域推荐网络的输出结果中。在单个患者的RPN算法输出中,单侧可能有40~80个中耳结构图,为下一个模块构建中耳特征块保留了较多的原始特征。

1.6 基于孪生网络的对比算法

在HRCT图像的成像过程中,患者头部姿势的变化会直接影响HRCT图像的扫描角度,致使目标选取的多个中耳特征结构(如水平半规管、鼓室窦口、内听道)未显示在同一张HRCT扫描图像中,从而导致待选择的中耳特征图混淆于所有的RPN输出结果中。对此,本研究选择基于孪生网络^[11]结构的算法来解决此类问题。孪生网络是用于计算同一领域下的图像相似度的方案。通过共享参数与结构的卷积神经网络构建两张图像的深层语义表征向量,并通过训练好的全连接网络来输出两张图的相似概率值。式(2)给出了相似度度量的标准。

$$D_w(X_1, X_2) = \|N_w(X_1) - N_w(X_2)\| \quad (2)$$

其中, X_1 和 X_2 表示两个输入的对比度图像, N_w 表示共享参数 w 的分析图像卷积网络, D_w 表示带参数 w 的两个图像之间的距离。最后,两个图像的相似性通过上述定义的距离来测量。

在该部分,模型将区域推荐网络所生成的特征

图与预设的10张特殊部位的中耳结构标准图进行相似度对比, 并按照对比结果构建患者的左、右中耳图像特征块, 左耳、右耳各5张。图2左下角显示了这个部分的结构细节。模型根据问题需求使用两种特殊模块。第一种是普通下采样块(图2中橙色块)。它由单个的最大池化、批量归一化和2D卷积操作组成。第二种是残差卷积块(图2中粉色块)。为了使构建的孪生网络拥有较高的图像泛化能力, 在残差卷积块中构建了更深的卷积操作, 并利用批量归一化和残差连接结构解决多层卷积中分布不稳定导致的梯度消失和梯度爆炸问题。图3中的A部分结构展示了单个残差卷积块的设计。最后孪生网络会将10张备选图像拼接成2个中耳特征数据块, 并输入到后续的3D-CNN中。

1.7 3D-CNN

3D-CNN注重通道之间的上下文语义, 进行3D层面的信息分析。3D-CNN在训练过程中, 层级网络中使用跨通道的滤波块来对输入数据进行3个方向的滑动式解析计算, 并学习得到符合损失最低的滤波参数, 从而达到分析图像块深层语义的目的。式(3)和式(4)为通过一次一组3D卷积后所生成的特征图的过程。

$$F^n = \{f_{ijk}^n\} \quad (3)$$

$$f_{ijk}^n = \sum_{k=1}^{c_1} \sum_{j=1}^{c_2} \sum_{i=1}^{c_3} f_{i+l, j+m, k+n}^{n-1} \times \sigma_{ijk} \quad (4)$$

其中, f^n 表示第 n 层的特征图块, 而 f_{ijk}^n 表示第 n 层特征图块上对应位置 i, j, k 的信息。 l, m, n 表示操作的起始位置。 c_1, c_2, c_3 分别表示卷积核在3D方向的大小; 在本文中 $c_1 = c_2 = c_3 = 3$ 。而 σ 则表示卷积核。模型最后经过3层交替的3D卷积块操作并经过全连接模块后, 会输出一个3D的张量, 代表样本对于三分类的置信度结果。

在3D-CNN输入前, 孪生网络部分会将提取的5张中耳特征图像拼接, 得到了一个尺寸为 $100 \times 100 \times 3 \times 5$ 的中耳特征图像数据块(默认输入的图像为彩色的RGB三通道图像)。如图2中所展示的3D-CNN部分, 笔者通过3次递进的3D卷积模块(卷积核尺寸为 $3 \times 3 \times 3$)与池化操作, 用原尺寸的特征图像块编码尺寸为 $4 \times 4 \times 1 \times 64$ 的特征块。

在3D-CNN中, 每个3D卷积块操作由2个3D卷积操作及2个LeakyRelu交替组成, 通过设定卷积过程中是否延拓控制图像的收缩过程。图3中的第2个结构为本文模型中的3D卷积模块。经过3轮采样后, 笔者将所获得的特征图重塑尺寸为1024的特征向量。并经过稠密全连接模块对提取出来的特征进行最后的分析。图3中的第3个结构为稠密全连接层的结构。

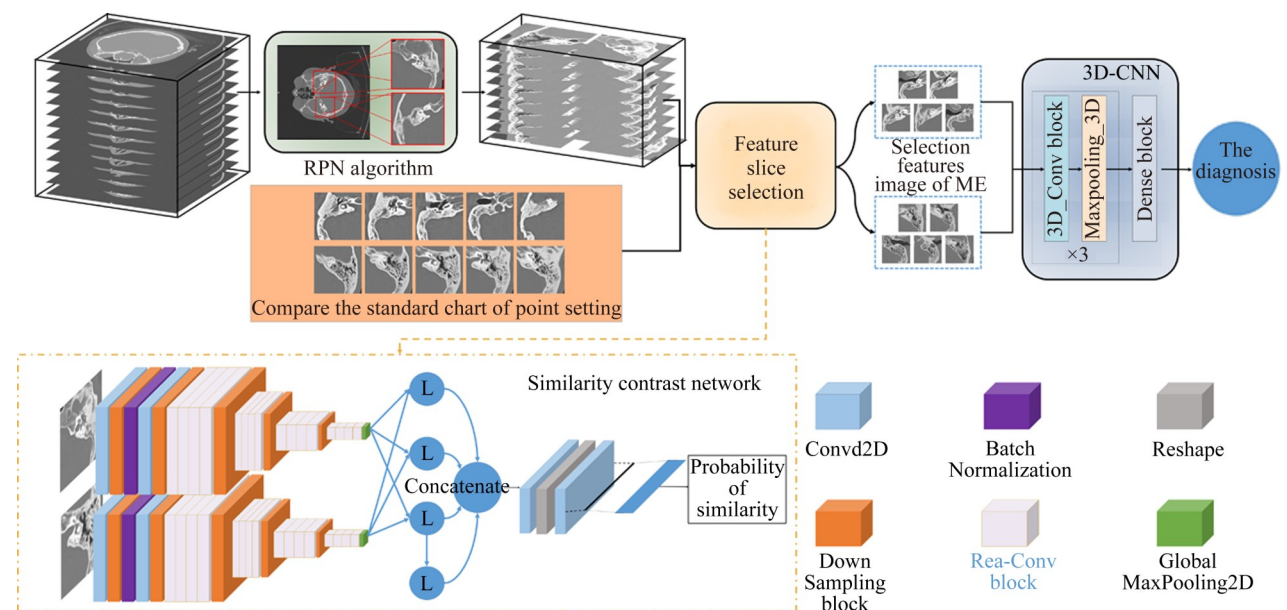


图2 模型的整体结构

Figure 2 Structure of model

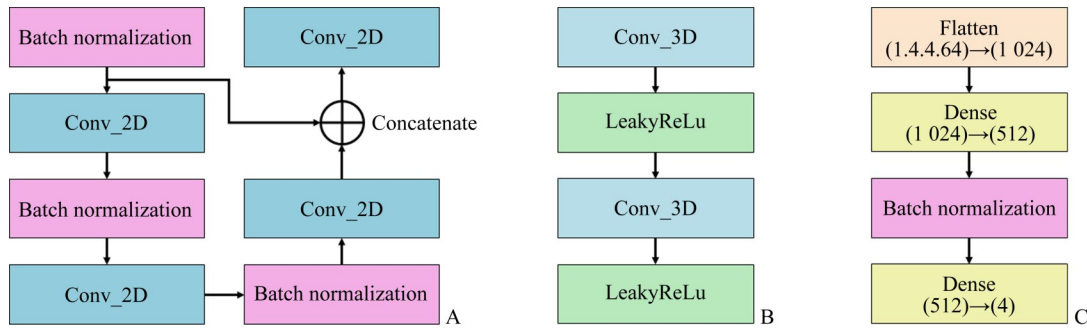


图3 模型内部分模块设置细节

Figure 3 Detail diagram of module setup in model

A: Residual connected sampling convolution module; B: 3D convolution module; C: 3D-CNN fully connected module

1.8 模型实施与训练细节

本实验将数据集划分为3个部分：训练集(75%)、验证集(10%)和测试集(15%)。图4为本研究的实验流程图。3个部分数据集均为实验数据集中按比例随机抽取构成，首先通过训练集得到模型的主要参数集，然后通过验证集对模型进一步微调，最后，利用训练好的模型对测试集图像进行测试，并输出测试结果。将预测结果与准确的标签进行比较，最终返回预测精度。

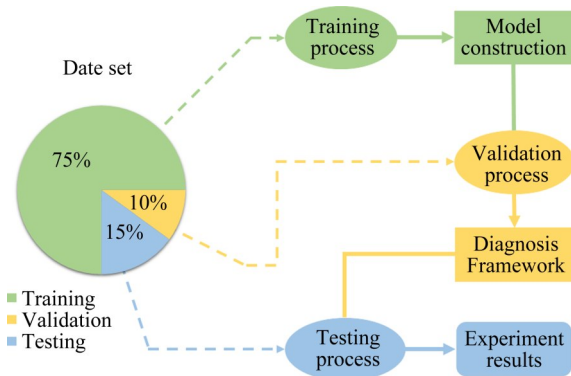


图4 实验流程示意图

Figure 4 Schematic diagram of experimental process

在构建不同的网络模块时，模型根据不同阶段的需求和目标设定相应的优化方法和损失函数。

在区域推荐网络中，将损失函数定义为 *dice-coefficient*，其主要度量网络切割出来的图像与训练图像的差异程度。方程如式(5)：

$$Dice = \frac{2 \times \sum_{ij} \min(x_{ij}, y_{ij})}{\sum_{ij} x_{ij} + \sum_{ij} y_{ij}} \quad (5)$$

其中， x_{ij}, y_{ij} 代表模型预测的感兴趣区域的像素灰度值和原始预测图的像素灰度值(黑色为背景、白色为划定区域)， i, j 为不同像素图像的坐标值， \min 代表选择括号中的最小值。

在基于孪生网络的特征图相似度对比网络中，使用的损失函数是 *binary-crossentropy*，其公式如下：

$$Loss = -\frac{1}{outputsize} \sum_{i=1}^{outputsize} y_i \times \log \hat{y}_i + (1 - y_i) \times \log(1 - \hat{y}_i) \quad (6)$$

在最后的3D-CNN中，笔者选择 *LeakyRelu* 作为激活函数，*Adam* 为优化器，交叉熵损失为损失函数。*LeakyRelu* 的计算公式如下：

$$LeakyRelu(x) = \begin{cases} x & x > 0 \\ ax & x \leq 0. \end{cases} \quad (7)$$

Adam 作为优化器，其依靠在反向传播时计算同层参数的梯度来构建梯度的一阶矩和二阶矩，使得学习率在训练过程中拥有自适应性^[12]，公式如下：

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (8)$$

其中 α 为学习率， \hat{m}_t 和 \hat{v}_t 为矫正偏差后的梯度均值和方差， ϵ 为防止分母为0的极小常数。

交叉熵损失函数用于评价当前训练得到的概率分布与真实分布之间的差异。交叉熵的值越小，两个概率分布越接近，公式如下：

$$Loss = - \sum_{i=1}^{outputsize} y_i \times \log \hat{y}_i \quad (9)$$

其中 \hat{y}_i 为预测输出值， y_i 为真实值。

本研究还对其他几种常见的卷积神经网络进行了特点分析与实验(表2)，结果显示相关参数的设置与训练本研究模型时提出的一致。

受试者工作特征(receiver operating characteristic, ROC)曲线常用于描述智能算法的适用性。在后续实验中，我们采用 ROC 曲线、曲线下面积(area under

the curve, AUC)值以及混淆矩阵的指标来评判模型的性能。本模型的第2部分采用ROC曲线描述了针对不同特定位置时孪生网络的泛用性能力。当AUC值越接近1时,说明该网络的泛化性越强。

表2 常见的卷积神经网络模型

Table 2 Common convolutional neural network models

名称	特点
ResNet ^[13]	引入残差链接,实现了更深层的分析
VGG16 ^[14]	经典的多阶段卷积神经网络
Inception-V3 ^[15]	利用分块卷积和 aggressive 正则提高模型效率
DenceNet ^[16]	引入稠密链接,使得网络易于训练
Shufflenet ^[17]	利用通道混排,减少了计算量并且提高了模型的稳定性

2 结果

2.1 数据预处理结果

经过高斯噪声分类别对原始数据集扩充后,MEC组数据分别经过了均值为0,方差为0.001~0.005的6倍数据增强,病例由原来的108侧耳扩增为648侧耳;正常组数据分别经过了均值为0,方差为0.001的2倍数据增强,病例由原来的268侧耳扩增为536侧耳,整体数据集由998侧耳扩增至1806侧耳。训练集1355侧耳,验证集180侧耳,测试集271

侧耳。

2.2 孪生网络实验结果与可视化

在孪生网络相似度对比实验中,将所有提取出来的中耳特征图与设定好的特定位置范例中耳特征图进行比较,然后输出对应位置计算相似程度最大的置信度对应的图像作为选择结果。图5为相似度对比的可视化过程,每个子图的上下一对图为一对比结果。结构相似的图像置信度为90%以上,而结构非明显相似的图像置信度只有70%左右(图5A);对比图像左右不一样时,相似程度的置信度为0(图5B)。本研究结果显示,ROC曲线结果显示孪生网络在该模型中的平均AUC为0.939(图6)。

2.3 模型最终诊断结果

经过前面的筛选之后,模型通过3D卷积网络最后输出的置信度来确定整体的分类结果。结果显示:最后诊断的准确度为96.5%,总体召回率为96.4%,F1得分为0.964,总体结果分类精度较高。

混淆矩阵分别对应不同的预测情况,以反映每一个样本的真实标签与被预测的情况,其中对角线上的元素代表被正确诊断的类别。结果显示:3D-CNN在分析HRCT连续数据上具有较好的适用性,仅有小部分CSOM图像被误诊为MEC。图7为分类结果的混淆矩阵。

经过模型前两个算法处理,测试集中的中耳特征块在最后3D-CNN判类算法上表现出高精度结果。3种判类结果的平均AUC值为0.983。图8为最后3D-CNN的ROC曲线图。

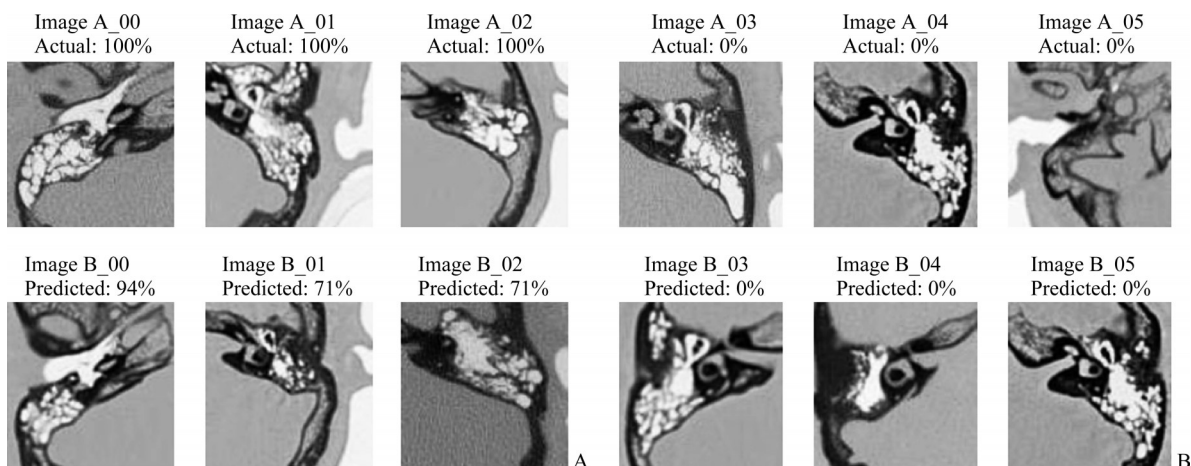


图5 孪生网络对比结果可视化

Figure 5 Visualization result of siamese networks

A: Comparative results of similar images; B: Comparative results of dissimilar images

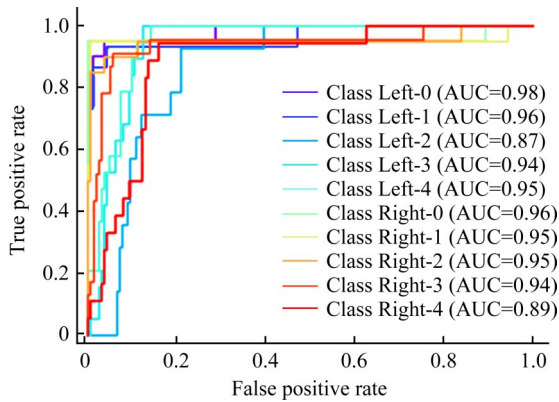


图6 孪生网络对比结果的受试者工作特征曲线图
Figure 6 Receiver operating characteristic curve of siamese network

Label	Normal	CSOM	MEC
Normal	98.3%		1.7%
CSOM		93.7%	6.3%
MEC	2.6%		97.4%
	Normal	CSOM	MEC

图7 模型诊断结果混淆矩阵
Figure 7 Confusion matrix of diagnosis results

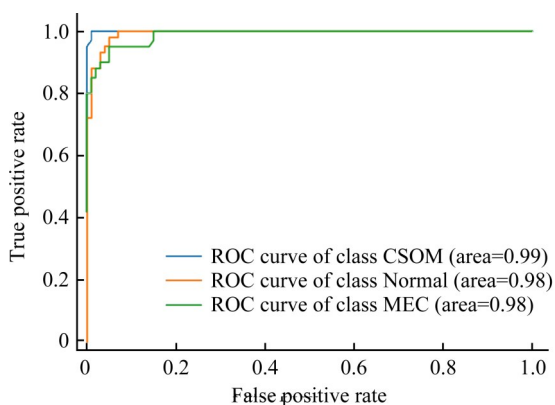


图8 模型诊断结果的受试者工作特征曲线图
Figure 8 Receiver operating characteristic curve of diagnosis results

2.4 与其他方法的对比实验结果

笔者针对一些常规的卷积神经网络方法、既往研究和本研究模型进行对比实验。在实验过程中，

为体现实验结果的可靠性，超参数的设定均与前述模型训练一致。在数据集方面，为了使实验具有对比性，笔者记录了每次划分数据集的随机数种子，在相同的随机数种子下进行对比试验，并得到精度的评价指标。表3为针对CSOM所进行的实验获得的准确度对比，结果显示：本研究提出的辅助诊断模型的准确度优于其他模型。

表3 几种卷积神经网络模型的训练结果

Table 3 Result of training by several convolutional neural network models

名称	准确度/%	召回率/%	F1得分
ResNet	80.6	92.8	0.863
VGG16	77.8	88.6	0.829
Inception-V3	77.4	88.2	0.827
DenceNet	80.6	89.5	0.848
Shufflenet	82.9	91.5	0.870
Pre-work	76.7	75.0	0.758
本研究的模型	96.5	96.4	0.964

3 讨论

在对中耳疾病的研究及认识过程中，病因及其临近器官的相互影响作用逐步明确，使得中耳疾病的治疗趋于系统化、规范化^[18-19]。CSOM的治疗原则为控制感染、通畅引流、清除病灶、恢复听力、消除病因；而MEC具有破坏周围骨质的特点，治疗原则为彻底清除病变组织、预防并发症，在可能的情况下重建中耳传音结构，故对MEC的早发现、早诊断、早治疗具有重要意义^[20-21]。

在中耳疾病诊疗过程中，颞骨HRCT能清晰地显示耳部及其邻近组织的精细解剖结构，对耳部先天畸形、各种中耳炎症及某些耳源性颅内并发症等具有较高的诊断价值，还可以帮助临床医生评估病变的程度，指导手术方案的制订等^[22-25]。CSOM与MEC可有相同的HRCT影像特征，可能会导致MEC的误诊或漏诊。Singh等^[26]采用术前HRCT影像诊断MEC，并与术后病理分析结果比对，结果显示术前HRCT诊断MEC的准确度为82%。减少疾病的误诊和漏诊是临床医师的工作目标，除临床经验的积累外，将患者信息与人工智能相结合，对疾病进行快速、准确的诊断是医学发展的大势所趋。

临床上，耳科疾病的诊断常需结合病史、体征、耳内窥镜检查、听力学检查、HRCT和MRI检查，其

中, 基于深度学习的卷积神经网络在医学图像的快速诊断中显示了其优越性。Hallac等^[27]基于卷积神经网络算法, 通过耳廓体征诊断耳廓畸形。Cha等^[28]构建的卷积神经网络模型在多种鼓膜、外耳道及中耳病变的诊断中取得了满意的效果, 分类准确度达93.67%。Nie等^[29]设计了新的卷积神经网络用于分析宽频声导抗图像报告, 以此辅助诊断耳硬化症。Heutink等^[30]使用多种卷积神经网络算法对颞骨HRCT数据进行分析, 分割和自动测量双耳重度或极重度患者耳蜗的大小及形态, 指导人工耳蜗的植入。Cho等^[31]应用CNN分割钆造影MRI内耳影像, 量化内淋巴囊积水, 对梅尼埃病的诊断及预后评估有较好的作用。以上案例均说明卷积神经网络与疾病诊疗的关系越来越密切, 且卷积神经网络具有较大的应用前景及发展空间。

既往相关的人工智能辅助诊断模型多数基于单一图像, 通过卷积神经网络来分析图像。而对于HRCT图像而言, 其断层扫描的方式使得自身上下图像存在关联性, 若单纯使用卷积神经网络就忽略了这一点。为了更好地解决这些问题, 本研究提出了一种中耳疾病人工智能辅助诊断模型, 模型中应用了3D-CNN方法, 使得整体模型能够分析图像集的上下关联信息, 使整体分析结果更合理、更准确。为进一步提高模型的诊断效率, 该模型通过在3D卷积判类网络前部署了串联式的区域推荐网络算法和孪生网络算法, 对输入图像集进行降噪和筛选处理。区域推荐网络算法用于从原始HRCT图像上获得中耳部分的特征图, 孪生网络算法用于自动选取5张中耳特征图像(包括上半规管平面、水平半规管平面、外耳道平面等)。通过这两个算法的部署, 整体模型实现了从100~300张高密度HRCT图像中自动选取5张中耳特征图像的功能, 并确保输入数据的简洁性与上下连续性。实验结果显示本模型具有高效性、可行性。基于区域推荐网络中得到的中耳特征图, 孪生网络算法对定义的10个特殊层面的平均AUC值为0.939。3D-CNN在最终诊断阶段的准确度为96.5%, 总体召回率为96.4%, F1评分为0.964。其中, CSOM的召回率为93.7%, MEC的召回率为97.4%。在最后分类的AUC平均值为0.985。总体而言, 该模型优于目前对CSOM和MEC进行智能诊断的其他模型, 且稳定性高, 更符合HRCT图像数据的连续性本质。而且, 该模型能够很好地区分3类不同的HRCT图像, 且整体准确度较高、适用性广; 但在错误分类情况方面, CSOM被误诊为MEC这一分类上测试病例较多。这可能是由于该部分数据的CSOM患者中耳炎症较重, 从图像上可观察到该部分病例的中

耳腔内被积液(或增生肉芽)填充, 在影像特点上可能与MEC类似, 造成了误判。在后续的研究中, 为了强化这方面的诊断, 可以加入新数据对模型进行强化训练, 加大真实病例的数据量, 使模型对多种情况进行全面地学习, 完善该模型的整体运行。

综上所述, 本研究提出的深度学习模型为耳鼻喉科人工智能的发展和推广提供了一个很好的方案, 有利于人工智能在智慧医疗系统中的稳定运行和云端部署。未来可以扩充数据集对该模型进行更新, 同时根据预测结果对预诊断确诊病例进行个体三维重建, 为医患沟通提供更好的媒介。

作者贡献声明: 粟日 实验操作, 论文设计、撰写与修改; 宋剑 论文设计与修改; 汪政 部分实验与数据分析; 卯爽、毛亦韬 数据采集; 吴学文、侯木舟 论文指导。所有作者阅读并同意最终的文本。

利益冲突声明: 作者声称无任何利益冲突。

参考文献

- [1] 张文韬. 非炎症鼓膜完整的传导性聋临床特征及耳内镜听骨链重建疗效研究[D]. 合肥: 安徽医科大学, 2021. ZHANG Wentao. Clinical characteristics and endoscopic ossiculoplasty of non-inflammatory conductive hearing loss with intact tympanic membrane[D]. Hefei: Anhui Medical University, 2021.
- [2] 魏兴梅, 陈彪, 崔丹默, 等. 分泌性中耳炎临床应用指南(2004版修订)[J]. 中国耳鼻咽喉头颈外科, 2016, 23(8): 454-472. <https://doi.org/10.16066/j.1672-7002.2016.08.006>. WEI Xingmei, CHEN Biao, CUI Danmo, et al. Clinical practice guideline: otitis media with effusion (update) [J]. Chinese Archives of Otolaryngology-Head and Neck Surgery, 2016, 23(8): 454-472. <https://doi.org/10.16066/j.1672-7002.2016.08.006>.
- [3] 朱富高, 孙美红, 华辉. 慢性化脓性中耳炎的听骨链病变及对听力的影响[J]. 临床耳鼻咽喉头颈外科杂志, 2008, 22(7): 312-314. <https://doi.org/10.3969/j.issn.1001-1781.2008.07.009>. ZHU Fugao, SUN Meihong, HUA Hui. Destruction of ossicular chain and its impact on hearing of patients with chronic suppurative otitis media[J]. Journal of Clinical Otorhinolaryngology Head and Neck Surgery, 2008, 22(7): 312-314. <https://doi.org/10.3969/j.issn.1001-1781.2008.07.009>.
- [4] 李陈, 王冰, 张韩, 等. 中耳胆脂瘤手术治疗进展[J]. 临床耳鼻咽喉头颈外科杂志, 2021, 35(10): 952-956. <https://doi.org/10.13201/j.issn.2096-7993.2021.10.020>. LI Chen, WANG Bing, ZHANG Han, et al. Advances in the surgical treatment of cholesteatoma of the middle ear[J]. Journal of Clinical Otorhinolaryngology Head and Neck

- Surgery, 2021, 35(10): 952-956. <https://doi.org/10.13201/j.issn.2096-7993.2021.10.020>.
- [5] 董蕊, 宋智慧, 徐姗姗, 等. 慢性化脓性中耳炎围术期抗菌药物预防应用临床路径对抗菌药物合理使用的影 响[J]. 实用药物与临床, 2021, 24(10): 944-946. <https://doi.org/10.14053/j.cnki.ppcr.202110018>.
- DONG Rui, SONG Zhihui, XU Shanshan, et al. Effect of clinical pathways on prophylactic use of antibiotics in the perioperative period of chronic suppurative otitis media[J]. Practical Pharmacy and Clinical Remedies, 2021, 24(10): 944-946. <https://doi.org/10.14053/j.cnki.ppcr.202110018>.
- [6] 孙晓卫, 张建基, 丁元萍, 等. 高分辨率CT对慢性化脓性中耳炎和胆脂瘤中耳炎软组织分型的诊断价值[J]. 中华耳鼻咽喉头颈外科杂志, 2011, 46(5): 388-392. <https://doi.org/10.3760/cma.j.issn.1673-0860.2011.05.009>.
- SUN Xiaowei, ZHANG Jianji, DING Yuanping, et al. Efficacy of high-resolution CT in differential diagnosis of chronic suppurative otitis media and cholesteatoma otitis media by soft-tissue shadows[J]. Chinese Journal of Otorhinolaryngology Head and Neck Surgery, 2011, 46(5): 388-392. <https://doi.org/10.3760/cma.j.issn.1673-0860.2011.05.009>.
- [7] 陆书昌, 范静平, 吕光宇. 慢性化脓性中耳炎的听骨破坏方式[J]. 临床耳鼻咽喉科杂志, 1991, 5(3): 147-149.
- LU Shuchang, FAN Jingping, LÜ Guangyu. Destruction of ossicles in chronic suppurative otitis media[J]. Journal of Clinical Otorhinolaryngology, 1991, 5(3): 147-149.
- [8] 李惠芬, 蒋向前, 李柱. 高斯滤波稳健性能的研究与改进[J]. 仪器仪表学报, 2004, 25(5): 633-637. <https://doi.org/10.19650/j.cnki.cjsi.2004.05.019>.
- LI Huifen, JIANG Xiangqian, LI Zhu. Study and improvement of robust performance of Gaussian filtering[J]. Chinese Journal of Scientific Instrument, 2004, 25(5): 633-637. <https://doi.org/10.19650/j.cnki.cjsi.2004.05.019>.
- [9] Ren SQ, He KM, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(6): 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>.
- [10] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]// Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969. NY: IEEE. <https://doi.org/10.1109/TPAMI.2018.2844175>.
- [11] Chopra S, Hadsell R, Lecun Y. Learning a similarity metric discriminatively, with application to face verification[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. June 20-25, 2005, San Diego, CA, USA. IEEE, 2005: 539-546. <https://doi.org/10.1109/CVPR.2005.202>.
- [12] Kingma DP, Ba J. Adam: A method for stochastic optimization [C/OL]. Proceedings of the international conference on learning representations. San Diego, USA. 2015[2021-08-12]. <https://doi.org/10.48550/arXiv.1412.6980>.
- [13] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 770-778. NY: IEEE. <https://doi.org/10.1109/CVPR.2016.90>.
- [14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. arXiv:1409.1556v6[cs.CV] (2015-04-10) [2021-03-28]. <https://doi.org/10.48550/arXiv.1409.1556>.
- [15] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 2818-2826. NY: IEEE. <https://doi.org/10.1109/CVPR.2016.308>.
- [16] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 2261-2269. <https://doi.org/10.1109/CVPR.2017.243>.
- [17] Zhang XY, Zhou XY, Lin MX, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 6848-6856. NY: IEEE. <https://doi.org/10.1109/CVPR.2018.00716>.
- [18] Harmes KM, Blackwood RA, Burrows HL, et al. Otitis media: diagnosis and treatment[J]. Am Fam Physician, 2013, 88(7): 435-440.
- [19] Shirai N, Preciado D. Otitis media: what is new? [J]. Curr Opin Otolaryngol Head Neck Surg, 2019, 27(6): 495-498. <https://doi.org/10.1097/MOO.0000000000000591>.
- [20] Schilder AGM, Marom T, Bhutta MF, et al. Panel 7: otitis media: treatment and complications[J]. Otolaryngol Head Neck Surg, 2017, 156(suppl 4): S88-S105. <https://doi.org/10.1177/0194599816633697>.
- [21] Wallis S, Atkinson H, Coatesworth AP. Chronic otitis media[J]. Postgrad Med, 2015, 127(4): 391-395. <https://doi.org/10.1080/00325481.2015.1027133>.
- [22] Baklaci D, Kuzucu I, Guler I, et al. Effect of mastoid bone pneumatization on the conformation and depth of the sinus tympani, a high-resolution computed tomography study[J]. Surg Radiol Anat, 2019, 41(8): 921-926. <https://doi.org/10.1007/s00276-019-02246-3>.
- [23] Feraco P, Piccinini S, Gagliardo C. Imaging of inner ear malformations: a primer for radiologists[J]. Radiol Med, 2021, 126(10): 1282-1295. <https://doi.org/10.1007/s11547-021-01387-z>.
- [24] Majeed J, Sudarshan Reddy L. Role of CT mastoids in the diagnosis and surgical management of chronic inflammatory ear diseases[J]. Indian J Otolaryngol Head Neck Surg, 2017, 69(1): 113-120. <https://doi.org/10.1007/s12070-016-1023-z>.
- [25] Seven H, Ozbal AE, Turgut S. Management of otogenic lateral sinus thrombosis[J]. Am J Otolaryngol, 2004, 25(5): 329-333. <https://doi.org/10.1016/j.amjoto.2004.04.005>.
- [26] Singh R, Rai R, Singh P, et al. High-resolution computed tomography (HRCT) in pediatric and adult patients with unsafe chronic suppurative otitis media (CSOM) and its surgical

- correlation[J]. J Family Med Prim Care, 2020, 9(8): 4067-4073. https://doi.org/10.4103/jfmpe.jfmpe_455_20.
- [27] Hallac RR, Lee J, Pressler M, et al. Identifying ear abnormality from 2D photographs using convolutional neural networks[J]. Sci Rep, 2019, 9(1): 18198. <https://doi.org/10.1038/s41598-019-54779-7>.
- [28] Cha D, Pae C, Seong SB, et al. Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database[J]. EBioMedicine, 2019, 45: 606-614. <https://doi.org/10.1016/j.ebiom.2019.06.050>.
- [29] Nie LX, Li C, Marzani F, et al. Classification of wideband tympanometry by deep transfer learning with data augmentation for automatic diagnosis of otosclerosis[J]. IEEE J Biomed Health Inform, 2022, 26(2): 888-897. <https://doi.org/10.1109/JBHI.2021.3093007>.
- [30] 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. <https://doi.org/10.1016/j.cmpb.2020.105387>.
- [31] Cho YS, Cho K, Park CJ, et al. Automated measurement of hydrops ratio from MRI in patients with Ménière's disease using CNN-based segmentation[J]. Sci Rep, 2020, 10(1): 7003. <https://doi.org/10.1038/s41598-020-63887-8>.

(本文编辑 陈丽文)

本文引用: 粟日, 宋剑, 汪政, 卯爽, 毛奔韬, 吴学文, 侯木舟. 基于3D卷积神经网络的中耳疾病高分辨率CT图像辅助分类诊断模型的应用[J]. 中南大学学报(医学版), 2022, 47(8): 1037-1048. DOI:10.11817/j.issn.1672-7347.2022.210704

Cite this article as: SU Ri, SONG Jian, WANG Zheng, MAO Shuang, MAO Yitao, WU Xuewen, HOU Muzhou. Application of high resolution computed tomography image assisted classification model of middle ear diseases based on 3D-convolutional neural network[J]. Journal of Central South University. Medical Science, 2022, 47(8): 1037-1048. DOI: 10.11817/j. issn. 1672-7347.2022. 210704