

基于VGG的慢性鼻窦炎分类模型研究*

张宁¹ 马瑞霞¹ 任海玲¹ 中学良¹ 贺娇¹ 赵玉桐¹ 杨凤霞¹ 刘铭¹ 王乐¹ 章峪侨² 曾芷灵²

[摘要] 目的:搭建基于VGG的慢性鼻窦炎计算机辅助诊断模型,并评价其效能。方法:①收集5 000帧已确诊的鼻窦CT图像,将其分为正常组1 000帧图像(其中,正常的上颌窦、额窦、筛窦、蝶窦影像图像各250帧)及异常组4 000帧图像(其中,上颌窦炎、额窦炎、筛窦炎、蝶窦炎影像图像各1 000帧),对图像进行大小归一化及分割预处理;②训练模型并对其进行仿真实验,分别得到正常组、蝶窦炎组、额窦炎组、筛窦炎组以及上颌窦炎组5个分类模型,从准确度、精确度、灵敏度、特异度、判读时间及ROC曲线下面积(AUC)6个维度,客观评价模型分类效能;③随机选取200帧图像,通过模型与低年资医师组、中年资医师组、高年资医师组分别阅片构成对比试验,结合临床通过以上评价指标客观评价模型的效能。结果:①仿真实验:整个模型的识别准确率为83.94%,精确率为89.52%,灵敏度为83.94%,特异度为95.99%,平均每帧图像判读时间为0.20s;蝶窦炎的AUC为0.865(95%CI 0.849~0.881),额窦炎的AUC为0.924(0.911~0.936),筛窦炎的AUC为0.895(0.880~0.909),上颌窦炎的AUC为0.974(0.967~0.982)。②对比实验:在识别准确度上,模型为84.52%,低年资医师组为78.5%、中年资医师组为80.5%、高年资医师组为83.5%;在识别精确度上,模型为85.67%,低年资医师组为79.72%,中年资医师组为82.67%,高年资医师组为83.66%;在识别灵敏度上,模型为84.52%,低年资医师组为78.50%,中年资医师组为80.50%,高年资医师组为83.50%;在识别特异度上,模型为96.58%,低年资医师组为94.63%,中年资医师组为95.13%,高年资医师组为95.88%;在耗时上,模型平均每帧图像为0.20s,低年资医师组平均每帧图像为2.35s,中年资医师组平均每帧图像为1.98s,高年资医师组平均每帧图像为2.19s。结论:本研究强调了基于深度学习的慢性鼻窦炎人工智能诊断模型分类诊断慢性鼻窦炎的可能性;基于深度学习的慢性鼻窦炎人工智能诊断模型分类性能好,具有较高的诊断效能。

[关键词] VGG;医学图像;慢性鼻窦炎;计算机辅助诊断

DOI:10.13201/j.issn.2096-7993.2024.07.013

[中图分类号] R765.21 **[文献标志码]** A

Research on the classification model of chronic sinusitis based on VGG

ZHANG Ning¹ MA Ruixia¹ REN Hailing¹ SHEN Xueliang¹ HE Jiao¹ZHAO Yutong¹ YANG Fengxia¹ LIU Ming¹ WANG Le¹ZHANG Yuqiao² ZENG Zhiling²

(¹Yinchuan First People's Hospital, Yinchuan, 750004, China; ²The Second Clinical College of Ningxia Medical University)

Corresponding author: MA Ruixia, E-mail: mrx2023111111@163.com

Abstract Objective: To build a VGG-based computer-aided diagnostic model for chronic sinusitis and evaluate its efficacy. **Methods:** ① A total of 5 000 frames of diagnosed sinus CT images were collected. The normal group consisted of 1 000 frames (250 frames each of maxillary sinus, frontal sinus, septal sinus, and pterygoid sinus), while the abnormal group consisted of 4 000 frames (1 000 frames each of maxillary sinusitis, frontal sinusitis, septal sinusitis, and pterygoid sinusitis). ② The models were trained and simulated to obtain five classification models for the normal group, the pteroid sinusitis group, the frontal sinusitis group, the septal sinusitis group and the maxillary sinusitis group, respectively. The classification efficacy of the models was evaluated objectively in six dimensions: accuracy, precision, sensitivity, specificity, interpretation time and area under the ROC curve (AUC). ③ Two hundred randomly selected images were read by the model with three groups of physicians (low, middle and high seniority) to constitute a comparative experiment. The efficacy of the model was objectively evaluated using the aforementioned evaluation indexes in conjunction with clinical analysis. **Results:** ① Simulation experiment: The overall recognition accuracy of the model is 83.94%, with a precision of 89.52%, sensitivity of 83.94%, specificity of 95.99%, and the average interpretation time of each frame is 0.2 s. The AUC for spher-

*基金项目:宁夏回族自治区重点研发计划项目(No:2023BEG02019)

¹银川市第一人民医院(耳鼻咽喉头颈外科医院)(银川,750004)

²宁夏医科大学第二临床学院

通信作者:马瑞霞, E-mail: mrx2023111111@163.com

noid sinusitis was 0.865(95%CI 0.849-0.881), for frontal sinusitis was 0.924(0.991-0.936), for ethmoidoid sinusitis was 0.895(0.880-0.909), and for maxillary sinusitis was 0.974(0.967-0.982). ② Comparison experiment: In terms of recognition accuracy, the model was 84.52%, while the low-seniority physicians group was 78.50%, the middle-seniority physicians group was 80.50%, and the seniority physicians group was 83.50%; In terms of recognition accuracy, the model was 85.67%, the low seniority physicians group was 79.72%, the middle seniority physicians group was 82.67%, and the high seniority physicians group was 83.66%. In terms of recognition sensitivity, the model was 84.52%, the low seniority group was 78.50%, the middle seniority group was 80.50%, and the high seniority group was 83.50%. In terms of recognition specificity, the model was 96.58%, the low-seniority physicians group was 94.63%, the middle-seniority physicians group was 95.13%, and the seniority physicians group was 95.88%. In terms of time consumption, the average image per frame of the model is 0.20 s, the average image per frame of the low-seniority physicians group is 2.35 s, the average image per frame of the middle-seniority physicians group is 1.98 s, and the average image per frame of the senior physicians group is 2.19 s. **Conclusion:** This study demonstrates the potential of a deep learning-based artificial intelligence diagnostic model for chronic sinusitis to classify and diagnose chronic sinusitis; the deep learning-based artificial intelligence diagnosis model for chronic sinusitis has good classification performance and high diagnostic efficacy.

Key words VGG; medical image; chronic rhinosinusitis; computer aided diagnosis

慢性鼻窦炎(chronic rhinosinusitis, CRS)是一种病程超过12周的鼻腔及鼻窦黏膜的慢性炎症性疾病,是耳鼻咽喉头颈外科的常见病、多发病以及难治性疾病^[1]。据文献报道^[2-4],中国人群CRS的总体患病率为8.0%,CRS影响了我国大约1.07亿人,因环境污染及其他因素的影响,该数据呈逐年上升趋势。鼻窦CT是诊断CRS的重要依据,随着医学影像技术的广泛应用,大量的影像数据给临床医生及放射科医生的工作带来了挑战,医生有可能会因为经验不足、读者疲劳、注意力分散等原因使一些疾病漏诊或者误诊^[5]。随着精准医疗的提出,传统的人工阅片方式已无法满足需求。

人工智能(artificial intelligence, AI),是研究使计算机模拟人类某些思维过程和智能行为的一门学科。深度学习(deep learning, DL)是机器学习领域中的一个新的研究领域,可从大量的原始数据中快速提取抽象的、复杂的深度特征,是一种多层次表示学习方法,由于其易学性、通用性以及高效性而适用于AI辅助医疗,近年来成为研究热点。卷积神经网络(convolutional neural network, CNN)模型是深度学习中应用最广泛的一种模型,在识别、分析、处理图像方面与人类视觉系统处理图像原理类似。VGGNet是牛津大学计算机视觉组和Google Deep Mind公司研究员一起研发的一种新的深度卷积神经网络^[6],它在AlexNet模型的基础上进一步优化,能够在处理图像时提取更多抽象的特征,由于网络结构的深化,VGGNet模型在进行图像特征提取时具有较强的学习能力。基于VGG模型的医学应用报道也相对较多,Hong等^[7]构建了一种基于VGG的儿童面部自动识别模型用于唐氏综合征的筛查,准确率可高达88.6%。Lin等^[8]基于VGG框架的卷积神经网络对肝癌进行分类,结果肝细胞癌分类的准确率超过90.0%。

虽然VGGNet算法在特征提取及识别等方面已经取得较多成果,但关于CRS的AI诊断相关研究在国内外的文献中报道相对较少。

本研究以鼻窦CT原始图像为研究对象,以VGG网络结构为架构,搭建基于深度学习算法的CRS-AI辅助诊断模型,对CRS进行分类诊断,评价基于深度学习算法建立的AI辅助诊断模型分类诊断CRS的效能,为后续的CRS-AI诊断提供基础模型。本研究也可依靠其强大的图像处理技术,有效提高医师阅片识别准确度,降低医师的阅片时间,提高其工作效率,降低劳动强度,提高不同医师之间的诊断一致性。

1 基础知识

VGG神经网络模型^[6],VGGNet是由Alexnet发展而来的,具有如下特点:①结构简单;②使用小卷积核(3×3)和连续多个卷积层,减少了模型参数的同时增加了非线性映射;③使用2×2的小池化核;④通道数更多、特征度更宽、层数更深;⑤网络测试阶段将训练阶段的3个全连接替换为3个卷积,使得测试得到的全卷积网络因为没有全连接的限制,因而可以接收任意宽或高为的输入。VGG-16的网络结构见图1,给出了各个深度的卷积层使用的卷积核大小以及通道的个数。其中,VGG-16是一个由16层组成的深度卷积神经网络,包括13个卷积层和3个全连接层,VGG-19是一个由19层组成的深度卷积神经网络,包括16个卷积层和3个全连接层。VGG-16的具体网络结构见图2。

2 基于VGG的4组CRS分类模型

2.1 鼻窦CT图像数据集

本研究使用的鼻窦CT图像包括蝶窦炎、额窦炎、筛窦炎、上颌窦炎4组鼻窦CT图像各1000帧,以及蝶窦、额窦、筛窦、上颌窦正常的4组鼻窦CT共1000帧。图3显示了鼻窦异常CT和鼻窦

正常 CT 图像比较。数据集样本较小,为避免过拟合发生,本研究将数据集按 8 : 2 拆分为训练集和测试集,蝶窦、额窦、筛窦和上颌窦训练集和测试集分别为 1 600 帧和 400 帧,并构造与其相对应的二进制标签 y ,其中正常标记为 0,异常标记为 1。

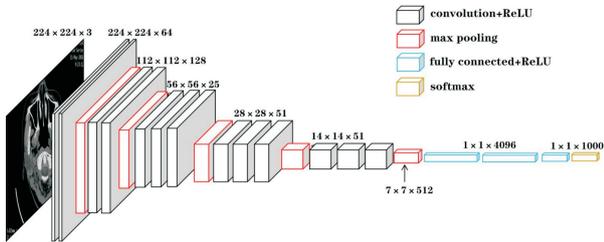


图 1 VGG-16 网络结构图

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weightlayers	11 weightlayers	13 weightlayers	16 weightlayers	16 weightlayers	19 weightlayers
Input (224×224 RGB image)					
Conv3-64	Conv3-64LRN	Conv3-64Conv3-64	Conv3-64Conv3-64	Conv3-64Conv3-64	Conv3-64Conv3-64
maxpool					
Conv3-128	Conv3-128	Conv3-128Conv3-128	Conv3-128Conv3-128	Conv3-128Conv3-128	Conv3-128Conv3-128
maxpool					
Conv3-256Conv3-256	Conv3-256Conv3-256	Conv3-256Conv3-256	Conv3-256Conv3-256Conv1-256	Conv3-256Conv3-256	Conv3-256Conv3-256Conv3-256
maxpool					
Conv3-512Conv3-512	Conv3-512Conv3-512	Conv3-512Conv3-512	Conv3-512Conv1-512	Conv3-512Conv3-512	Conv3-512Conv3-512Conv3-512
maxpool					
Conv3-512Conv3-512	Conv3-512Conv3-512	Conv3-512Conv3-512	Conv3-512Conv3-512Conv1-512	Conv3-512Conv3-512	Conv3-512Conv3-512Conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
Soft-max					

图 2 VGG 网络结构图



以上图均为水平位软组织窗鼻窦 CT。a~d: 双侧上颌窦、蝶窦、额窦、筛窦正常 CT; e: 右侧上颌窦腔内密度增高影, 左侧上颌窦腔内黏膜增厚; f: 双侧筛窦、蝶窦密度增高影; g: 双侧额窦腔内密度增高影; h: 双侧筛窦腔内密度增高影, 双侧蝶窦腔内正常。

图 3 4 组鼻窦炎和鼻窦正常 CT 图像比较

2.2 图像预处理

图像预处理是模型训练前的关键环节,影响模型最终效能好坏。本研究首先对图像进行缩放,调整图像整体大小并保留主要特征,归一化处理为大小为 $224\text{ cm} \times 224\text{ cm} \times 3\text{ cm}$;其次基于线性变换来提高模型对 CT 图像位置和方向的敏感性;最后对处理后的图像进行对比度增强,以提高模型对鼻窦 CT 图像的识别能力。

2.3 模型构建

本研究基于 VGG,使用 3×3 大小的卷积核进行卷积;使用 2×2 最大值池化进行降采样;循环 4 次卷积和降采样操作进入到全连接层,将包含有局部信息的特征图(包括特征图的高、宽、通道数)全

部映射到 4 096 维度。其网络结构设计规则为: ①输入: $224\text{ cm} \times 224\text{ cm} \times 3\text{ cm}$ 的通道图像进入卷积层。②图像预处理:对图像做均值预处理,每个像素中减去在训练集上计算的 RGB 均值,使得数据被强制地拉回至方差为 1、平均值为 0 的标准正态分布。③卷积层:网络使用连续的小卷积核 (3×3) 做连续卷积,卷积的固定步长为 1,并在图像的边缘填充 1 个像素,这卷积后保持图像的分辨率不变。④池化层:连续的卷积层会接着一个池化层,降低图像的分辨率。空间池化由 5 个最大池化层进行,在 2×2 像素窗口上进行最大池化,步长为 2。⑤进入全连接层,将包含有局部信息的特征图(包括特征图的高、宽、通道数)全部映射到 4 096

维度。⑥输出:卷积层后,接着的是 3 个全连接层,前两个每个都有 4 096 个通道。

3 模型仿真实验及对比实验

3.1 实验环境

软件环境: Windows10 操作系统, MatlabR2021a; 硬件环境: Intel (R) Core (TM) i5-7200UCPU@2.50 G Hz 2.70 G Hz,4.0 GB 内存,500 GB 硬盘。

3.2 实验结果及分析

3.2.1 模型实验结果分析 将鼻窦 CT 影像图像分为正常鼻窦 CT 及异常鼻窦 CT(上颌窦炎、额窦炎、筛窦炎、蝶窦炎),在训练阶段完成后,用测试数据集进行测试,得到混淆矩阵。混淆矩阵结果见表 1。

表 1 基于 VGG 的 4 组 CRS 分类模型混淆矩阵 帧

类型	蝶窦炎	额窦炎	筛窦炎	上颌窦炎	正常
蝶窦炎	678	30	169	42	81
额窦炎	0	820	2	2	176
筛窦炎	38	53	769	14	126
上颌窦炎	10	7	0	946	37
正常	1	0	10	5	984

基于混淆矩阵,可通过准确度、精确度、灵敏度、特异度、判读时间及 ROC 曲线 6 个指标对本研究提出的基于 VGG 的 CRS-AI 辅助诊断模型进行有效性评价。其中准确度、精确度、灵敏度、特异

度、判读时间及 AUC 评价,结果见表 2,ROC 曲线见图 4。

表 2 准确率、精确率、灵敏度、特异度、判读时间评价结果

类型	ACC /%	Precision /%	SEN /%	SPE /%	时间/s
蝶窦炎		93.26	67.80	98.78	
额窦炎		90.11	82.00	97.75	
筛窦炎	83.94	80.95	76.90	95.48	0.20
上颌窦炎		93.76	94.60	98.43	
正常		70.09	98.40	89.50	

从临床角度分析灵敏度和特异度。就该模型对 4 组的精确度而言:蝶窦炎、额窦炎、筛窦炎、上颌窦炎均具有较高的精确度,说明该模型分类效果较好,正常的精确度较低,考虑与 4 组鼻窦均正常视为正常,故精确度较低。就该模型对于 4 组鼻窦的特异度而言:特异度均较高,表明该模型具有较高的 CRS 检出率,亦 CRS 的误诊率较低。就该模型对于 4 组 CRS 的灵敏度而言:上颌窦炎、额窦炎的灵敏度相对于蝶窦炎和筛窦炎高,分析其可能原因是:相对于上颌窦、额窦在水平位 CT 图像中呈独立位置,而蝶窦、筛窦在水平位 CT 图像中位置重合较多,因此灵敏度较低,漏诊率较高。该模型具有较高的 CRS 检出率,但就蝶窦炎和筛窦炎判断容易出现混淆。

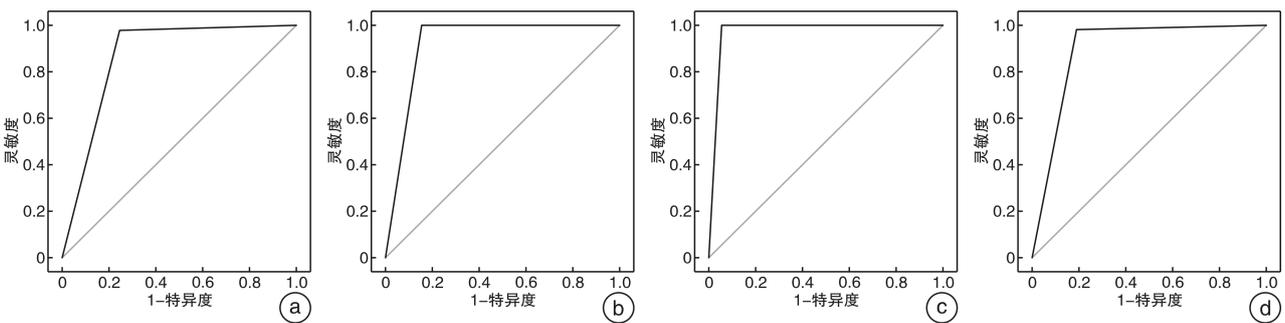


图 4 基于模型识别 4 组鼻窦炎的 ROC 曲线

基于 ROC 曲线,蝶窦炎 AUC 为 0.865(95% CI 0.849~0.881),额窦炎 AUC 为 0.924(95% CI 0.911~0.936),筛窦炎 AUC 为 0.895(95% CI 0.880~0.909),上颌窦炎 AUC 为 0.974(95% CI 0.967~0.982),依次见图 4a~d,可见 4 组鼻窦炎的 AUC 均>0.8,证明该分类器的分类性能较好。

3.2.2 对比试验结果分析 为科学评价本模型的效能,设立低年资医师组(2 名住院医师)、中年资医师组(2 名主治医师)、高年资医师组(1 名副主任医师+1 名主任医师)构造对比试验,分别对 200

帧鼻窦 CT(蝶窦炎 40 帧、额窦炎 40 帧、筛窦炎 40 帧、上颌窦炎 40 帧、鼻窦正常 40 帧)进行分类判别,得到 3 个年资组及模型分类混淆矩阵,见表 3。每组 2 名医师有不同意见时,则经共同商讨取得一致。所有阅片医师均记录每帧图像的阅片判读时间。

基于 3 个年资组及模型分类混淆矩阵,可分别得出低年资医师组、中年资医师组、高年资医师组、模型的阅片分类准确度、精确度、灵敏度、特异度、判读时间。低年资医师阅片评价结果见表 4、

中年资医师阅片评价结果见表 5、高年资医师阅片评价结果见表 6、模型阅片评价结果见表 7,3 个年资组及模型的分类结果比较见表 8。

表 3 不同年资医师组阅片混淆矩阵

组别	蝶窦炎	额窦炎	筛窦炎	上颌窦炎	正常
低年资组					
蝶窦炎	22	0	11	5	2
额窦炎	2	32	3	2	1
筛窦炎	4	1	33	1	1
上颌窦炎	1	1	2	35	1
正常	1	0	4	0	35
中年资组					
蝶窦炎	23	0	10	7	0
额窦炎	2	34	4	0	0
筛窦炎	2	4	34	0	0
上颌窦炎	0	0	3	37	0
正常	1	1	4	1	33
高年资组					
蝶窦炎	28	0	7	3	2
额窦炎	2	37	1	0	0
筛窦炎	2	4	31	1	2
上颌窦炎	0	0	0	35	5
正常	1	1	2	0	36
模型组					
蝶窦炎	27	1	7	2	3
额窦炎	0	33	0	0	7
筛窦炎	2	2	31	0	5
上颌窦炎	0	0	0	39	1
正常	0	0	1	0	39

表 4 低年资医师组阅片评价

类型	ACC/%	Precision /%	SEN/%	SPE/%	时间/s
蝶窦炎		73.33	55.00	95.00	
额窦炎		94.12	80.00	98.75	
筛窦炎	78.50	62.26	82.50	87.50	2.35
上颌窦炎		81.40	87.50	95.00	
正常		87.50	87.50	96.88	
AVE	78.50	79.72	78.50	94.63	2.35

表 5 中年资医师组阅片评价

类型	ACC/%	Precision /%	SEN/%	SPE/%	时间/s
蝶窦炎		82.14	57.50	96.88	
额窦炎		87.18	85.00	96.88	
筛窦炎	80.50	61.82	85.00	86.88	1.98
上颌窦炎		82.22	92.50	95.00	
正常		100.00	82.50	100.00	
AVE	80.50	82.67	80.50	95.13	1.98

表 6 高年资医师组阅片评价

类型	ACC/%	Precision /%	SEN/%	SPE/%	时间/s
蝶窦		84.85	70.00	96.88	
额窦		88.10	92.50	96.88	
筛窦	83.50	75.61	77.50	93.75	2.19
上颌窦		89.74	87.50	97.50	
正常		80.00	90.00	94.38	
AVE	83.50	83.66	83.50	95.88	2.19

表 7 模型阅片评价

类型	ACC/%	Precision /%	SEN/%	SPE/%	时间/s
蝶窦		93.26	67.80	99.53	
额窦		90.11	82.00	98.50	
筛窦	84.52	79.69	76.90	96.20	0.2
上颌窦		95.21	97.50	98.80	
正常		70.09	98.40	89.88	
AVE	84.52	85.67	84.52	96.58	0.2

表 8 3 个年资医师组及模型的分类 AVE 结果比较

组别	ACC /%	precision /%	SEN /%	SPE /%	时间/s
模型组	84.52	85.67	84.52	96.58	0.20
低年资组	78.50	79.72	78.50	94.63	2.35
中年资组	80.50	82.67	80.50	95.13	1.98
高年资组	83.50	83.66	83.50	95.88	2.19

由此可见:①在识别准确度上,模型识别准确度为 84.52%,较低年资医师组的准确度 78.5%、中年资医师组的 80.5%及高年资医师组的 83.5%均高;②在识别精确度上,模型识别精确度为 85.67%,较低年资医师组的精确度 79.72%、中年资医师组的精确度 82.67%及高年资医师组的精确度 83.66%均高;③在识别灵敏度上,模型识别灵敏度为 84.52%,较低年资医师组的灵敏度 78.50%、中年资医师组的灵敏度 80.50%及高年资医师组的灵敏度 83.50%均高;④在识别特异度上,模型识别特异度为 96.58%,较低年资医师组的特异度 94.63%、中年资医师组的特异度 95.13%及高年资医师组的特异度 95.88%均高。其在识别蝶窦炎的灵敏度及特异度上优于不同年资医师组;⑤在耗时上,模型每张阅片时间 0.20 s,显著低于低年资医师组平均每帧图像 2.35 s、中年资医师组平均每帧 1.98 s 和高年资医师组平均每帧 2.19 s。

4 讨论

CRS 是耳鼻咽喉科学的常见病及多发病,鼻窦 CT 作为诊断 CRS 的重要参考依据,往往是临床医

师及放射科医师根据临床经验及肉眼去判断,容易误诊,漏诊。随着计算机技术的进步,图像识别变得越来越容易,已有学者将其应用于CRS的辅助诊断。2018年,Chowdhury等^[9]探讨了使用CNN从鼻窦CT中自动识别临床相关信息的可行性,对239例CRS患者的鼻窦冠状位CT图像进行OMC的分类,在将OMC分类为“开放”或“阻塞”方面达到了85%的准确度。这是第一个基于深度学习算法来识别鼻窦CT的神经网络模型。同年,韩国学者Kim等^[10]通过比较深度学习算法与放射科医师在鼻窦X线片上诊断上颌窦炎的诊断性能。在2组测试中,深度学习算法的AUC均显著高于放射科医生。在灵敏度和特异度方面,深度学习算法与放射科医生相当。

目前尚未发现有研究对4组鼻窦炎进行分类诊断,本研究在此方面填上了空白,首次探讨了基于深度学习算法对4组鼻窦炎进行分类诊断的可行性。本研究中整个模型的识别准确度达到了83.94%,精确度89.52%,灵敏度83.94%,特异度95.99%,平均每帧图像耗时0.20s。其中,蝶窦炎与筛窦炎的识别灵敏度较低。分析原因:由于CNN技术的现状,目前只能分析单个二维CT图像,且鼻窦的解剖结构复杂,位置邻近,这给应用AI模型来分类检测蝶窦与筛窦带来困难和挑战,这可能是导致筛窦炎和蝶窦炎识别灵敏度低的主要原因。然而,即使2组鼻窦炎的灵敏度较低,但其识别特异度均较高,也就是说,当模型分类出是哪组鼻窦时,即能正确诊断为鼻窦炎的概率也高,误诊率较低。鉴于单个二维图像无法充分捕捉鼻窦CT的复杂性,还存在关于每个CT的“基本真相”的潜在分歧。本研究通过设定“正常”和“4组鼻窦炎”明确的先验规则,并由接受专业培训的鼻科医生结合临床症状及鼻内镜检查对扫描结果进行评估,部分缓解了这一问题。针对上述蝶窦与筛窦位置易混淆的问题,在后续研究中,拟通过对各个鼻窦区域进行标注,加大样本量继续训练模型,提高AI模型对蝶窦和筛窦的空间识别能力,提高灵敏度。4组鼻窦炎的AUC均 >0.8 ,该模型表现出较好的分类性能。

VGG网络是一种流行的CNN架构,其特点是,它只使用 3×3 的卷积层,这些卷积层彼此叠加,深度不断增加。深度的增加和核的缩小可以减小网络参数,从而提高拟合能力和临床应用的广泛性。该网络已广泛应用于计算机视觉领域。最近,VGG的医学应用已被广泛报道,相关工作涉及肿瘤特性识别^[11]、医学图像数据疾病分期^[8]、皮肤损伤分类^[12]等领域。本研究基于VGG架构搭建的4组慢性鼻窦炎分类诊断模型在与不同年资医师比较中也表现出好效能:①在识别准确度、精确度、

灵敏度、特异度上,模型的4个指标略高于高年资医师组,高于中年资医师组及低年资医师组。尤其在识别蝶窦炎的灵敏度及特异度上优于不同年资医师组。模型整体表现出的诊断水平可与高年资医师相当,可在一定程度上辅助低中年资医师进行诊断,提高不同医师之间的诊断一致性。不同年资医师在识别蝶窦炎时易误诊为筛窦炎,分析原因考虑提供的图像为单个二维图像有关,在临床实践中,应连续上下多层面、多角度观察各鼻窦位置,有助于将上述鼻窦进行鉴别。②在耗时上,模型每张阅片时间0.20s,显著低于不同年资医师组,人工结合AI模型时,有望提高诊断效率,降低医师的压力。在与临床医师比较试验中,再次验证了基于VGG的AI在CRS辅助诊断模型的可行性。本研究基于VGG初步建立了检测4组鼻窦炎的AI模型,这是将AI技术应用于分类诊断CRS的新尝试。

本研究也存在一些局限性:①有限的数据集。首先,本研究只收集了正常鼻窦CT及CRS-CT,收集的影像图像不包括鼻窦的其他疾病,故本研究对于鼻窦其他疾病诊断具有局限性,因此为了在未来利用基于AI的辅助软件,后续研究会收集更多的其他鼻腔鼻窦疾病病例与之鉴别,有利于提高模型性能;其次,虽然与许多鼻科学研究比较,该数据集的大小适中,但与通常用于训练计算机视觉算法的数百万个数据点比较,它相形见绌。②本研究只采用了水平位鼻窦CT评估黏膜增厚或窦腔密度增高影,未收集冠状位及矢状位对鼻窦CT进行综合判断,后续研究若能将多个方位结合,效果会更好。③考虑未成年患者鼻窦未发育完全,筛选的鼻窦CT均 >18 岁,不适用于辅助诊断儿童及青少年患者CRS。④本研究不包括其他医疗中心的外部测试数据集。就CT设备而言,与其他医疗成像设备相比,性能差异相对较小,具体取决于制造商或型号。

本研究强调了基于深度学习的CRS-AI诊断模型分类诊断CRS的可能性;基于深度学习的CRS-AI诊断模型分类性能好,具有较高的诊断效能;基于深度学习的CRS-AI诊断模型在阅片耗时上,显著低于不同年资医师。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 中华耳鼻咽喉头颈外科杂志编辑委员会鼻科组,中华医学会耳鼻咽喉头颈外科学分会鼻科学组.中国慢性鼻窦炎诊断和治疗指南(2018)[J].中华耳鼻咽喉头颈外科杂志,2019,54(2):81-100.
- [2] Liu Z,Chen J,Cheng L,et al.Chinese society of allergy and Chinese society of otorhinolaryngology-head and neck surgery guideline for chronic rhinosinusitis [J].Allergy Asthma Immunol Res,2020,12(2):176-

- 237.
- [3] 姚和梅,刘领波,李红英,等.慢性鼻窦炎及鼻息肉与性别和年龄的关系[J].中国耳鼻咽喉头颈外科,2005(3):183-184.
- [4] 司马宇彤,赵妍,矫健,等.以慢性鼻窦炎内在型为导向的临床治疗选择[J].临床耳鼻咽喉头颈外科杂志,2023,37(11):902-908.
- [5] 赵愉,王得旭,顾力栩.人工智能技术在计算机辅助诊断领域的发展新趋势[J].中国科学:生命科学,2020,50(11):1321-1334.
- [6] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv E Prints,2014;arXiv:1409.1556.
- [7] Hong D, Zheng YY, Xin Y, et al. Genetic syndromes screening by facial recognition technology: VGG-16 screening model construction and evaluation[J]. Orphanet J Rare Dis,2021,16(1):344.
- [8] Lin HX, Wei C, Wang GX, et al. Automated classification of hepatocellular carcinoma differentiation using multiphoton microscopy and deep learning[J]. J Bio-
photonics,2019,12(7):e201800435.
- [9] Chowdhury NI, Smith TL, Chandra RK, et al. Automated classification of osteomeatal complex inflammation on computed tomography using convolutional neural networks[J]. Int Forum Allergy Rhinol,2019,9(1):46-52.
- [10] Kim Y, Lee KJ, Sunwoo L, et al. Deep learning in diagnosis of maxillary sinusitis using conventional radiography[J]. Invest Radiol,2019,54(1):7-15.
- [11] Khan HA, Jue W, Mushtaq M, et al. Brain tumor classification in MRI image using convolutional neural network[J]. Math Biosci Eng, 2020, 17(5): 6203-6216.
- [12] Romero-Lopez A, Giro-i-Nieto X, Burdick J, et al. Skin lesion classification from dermoscopic images using deep learning techniques[C]//Biomedical Engineering. Innsbruck, Austria. Calgary, AB, Canada: ACT-APRESS,2017.

(收稿日期:2023-12-22)

(上接第 623 页)

- [19] Kim SH, Choi BY, Park J, et al. Maternal and placental factors associated with congenital hearing loss in very preterm neonates[J]. Pediatr Neonatol,2017,58(3):236-244.
- [20] Waters TP, Silva N, Denney JM, et al. Neonatal hearing assessment in very low birth weight infants exposed to antenatal steroids[J]. J Perinatol,2008,28(1):67-70.
- [21] 孙鹏程,梁勇,谭曼玲,等.84例母体产前应用地塞米松新生儿畸变产物耳声发射结果初步分析[J].中国听力语言康复科学杂志,2015,13(4):257-261.
- [22] 盘琳琳,翟丰,陈洁.新生儿缺氧对听觉系统损伤的临床和基础研究进展[J].中国眼耳鼻喉科杂志,2022,22(6):647-650.
- [23] 陈爱平,王海波.耳后注射糖皮质激素治疗突发性聋的研究进展[J].临床耳鼻咽喉头颈外科杂志,2016,30(19):1572-1575.
- [24] 王慧,曹丽芳,张雪峰.产前糖皮质激素给药时机对早产儿结局的影响[J].中国小儿急救医学,2021,28(12):1094-1098.
- [25] Church MW, Adams BR, Anumba JI, et al. Repeated antenatal corticosteroid treatments adversely affect neural transmission time and auditory thresholds in laboratory rats[J]. Neurotoxicol Teratol,2012,34(1):196-205.
- [26] 许淑静,陈国洪,孙慧清,等.不同程度支气管肺发育不良早产儿神经系统发育对比研究[J].中华实用诊断与治疗杂志,2019,33(2):139-142.

(收稿日期:2023-11-27 修回日期:2024-01-30)