

DOI:10.3969/j.issn.1005-202X.2023.08.021

医学人工智能

## 基于深度卷积神经网络的上消化道内镜解剖分类模型构建

许郭婷<sup>1,2</sup>, 吴爱荣<sup>1,2</sup>, 林嘉希<sup>1,2</sup>, 高欣<sup>1,2</sup>, 周鑫<sup>3</sup>, 顾慧媛<sup>1,2</sup>, 许春芳<sup>1,2</sup>, 朱锦舟<sup>1,2</sup>

1. 苏州大学附属第一医院消化内科, 江苏 苏州 215000; 2. 苏州市消化病临床医学中心, 江苏 苏州 215000; 3. 江苏大学附属金坛医院消化内科, 江苏 常州 213200

**【摘要】目的:**利用深度卷积神经网络构建上消化道内镜解剖分类模型。**方法:**收集苏州大学附属第一医院消化内镜中心4183张胃镜图片,按照8:2的比例随机分为训练集和验证集;同时收集江苏大学附属金坛医院270张胃镜图片作为测试集。以上图片标注上消化道解剖位置(包括食管、贲门、胃底、胃体、胃角、胃窦、幽门、十二指肠球部及降部)。选择ImageNet数据库中预训练的Xception、NASNet Large(NASNet)和ResNet50V2(ResNet)3个深度卷积神经网络,在训练集及验证集中训练,构建上消化道图片解剖部位分类模型。使用梯度加权分类激活映射对模型的分类结果进行可视化解释。在验证集和测试集中评价模型分类能力。**结果:**成功构建了基于深度卷积神经网络的上消化道内镜解剖分类的3个模型,各模型均具备较高的分类能力。在验证集中,平均分类准确性为0.980,平均分类召回率为0.894,平均分类精确度为0.920;其中,ResNet模型表现最优,其分类准确性(0.982)、分类召回率(0.905)和分类精确度(0.933)最高。在测试集中,平均分类准确性为0.988,平均分类召回率为0.942,平均分类精确度为0.950;其中,NASNet模型表现最优,其分类准确性(0.992)、分类召回率(0.959)和分类精确度(0.970)最高。梯度加权分类激活映射以热力图形式对模型分类结果提供可视化解释。**结论:**利用深度卷积神经网络,构建的上消化道内镜解剖分类模型具有较好的分类能力。

**【关键词】**上消化道;胃镜;解剖定位;深度卷积神经网络;模型构建

**【中图分类号】**R318

**【文献标志码】**A

**【文章编号】**1005-202X(2023)08-1051-06

## Development of anatomical classification models for upper gastrointestinal endoscopy based on deep convolutional neural networks

XU Guoting<sup>1,2</sup>, WU Airong<sup>1,2</sup>, LIN Jiaxi<sup>1,2</sup>, GAO Xin<sup>1,2</sup>, ZHOU Xin<sup>3</sup>, GU Huiyuan<sup>1,2</sup>, XU Chunfang<sup>1,2</sup>, ZHU Jinzhou<sup>1,2</sup>

1. Department of Gastroenterology, the First Affiliated Hospital of Soochow University, Suzhou 215000, China; 2. Suzhou Clinical Center of Digestive Disease, Suzhou 215000, China; 3. Department of Gastroenterology, Jintan Hospital Affiliated to Jiangsu University, Changzhou 213200, China

**Abstract: Objective** To develop anatomical classification models for upper gastrointestinal endoscopy using deep convolutional neural networks. **Methods:** A total of 4 183 gastroscopic images collected from the Gastrointestinal Endoscopy Center of the First Affiliated Hospital of Soochow University were randomly divided into training set and validation set at a ratio of 8:2, while 270 gastroscopic images from Jintan Hospital Affiliated to Jiangsu University were collected as the test set. The anatomical structures (esophagus, cardia, gastric fundus, gastric body, gastric angle, gastric antrum, pylorus, duodenal bulb and descending) were labeled in the gastroscopic images. Three deep convolutional neural networks, namely Xception, NASNet Large (NASNet) and ResNet50V2 (ResNet), which had been pre-trained in ImageNet database, were trained in training set and validation set for constructing the anatomical classification models for upper gastrointestinal endoscopy. The gradient-weighted class activation mapping was used to visually interpret the classification results of the models, and the classification abilities of the models were evaluated in validation set and test set. **Results** Three anatomical classification models for upper gastrointestinal endoscopy based on deep convolutional neural network were successfully constructed. All models had high classification ability. In the validation set, the average classification accuracy, recall and precision were 0.980, 0.894 and 0.920, respectively. Among them, ResNet model performed best, with the highest classification accuracy

**【收稿日期】**2023-01-19

**【基金项目】**国家自然科学基金(82000540);苏州市科技计划项目(SKY2021038);苏州市科教兴卫项目(KJXW2019001)

**【作者简介】**许郭婷,在读研究生,E-mail: gtxu@stu.suda.edu.cn

**【通信作者】**朱锦舟,博士,副主任医师,硕士生导师,E-mail: jzzhu@zju.edu.cn

(0.982), classification recall (0.905) and classification precision (0.933). In the test set, the average classification accuracy, recall and precision were 0.988, 0.942 and 0.950, respectively. Among them, NASNet model performed best, with the highest classification accuracy (0.992), classification recall (0.959) and classification precision (0.970). The gradient-weighted class activation mapping provides a visual interpretation of the model classification results in the form of thermal map.

**Conclusion:** The anatomical classification model developed by deep convolutional neural network for upper gastrointestinal endoscopy has preferable classification ability.

**Keywords** upper gastrointestinal tract; gastroscope; anatomical location; deep convolutional neural network; model construction

前言

目前上消化道疾病的诊断主要依靠消化内镜检查<sup>[1]</sup>。随着放大内镜、双聚焦内镜、色素内镜、窄带成像以及智能分光比色等全新技术不断投入临床运用,极大提高了消化道肿瘤及病变的诊断水平<sup>[2]</sup>。内镜在上消化道疾病的筛查与诊断中发挥越来越重要的作用,但内镜诊断的准确性仍然依赖于操控者熟练的技能与丰富的经验,内镜医师需要进行专门的培训以及大量的操作经验<sup>[3-4]</sup>。随着上消化道疾病患者数量增加,内镜医生需要分析的数据量也越来越大。基于人工智能技术的高效智能诊断系统可以减轻医生负担,辅助医生诊断。近年来,基于深度学习的特征提取方法在各种图像处理领域表现出最佳性能,被广泛应用于医学领域,在放射学、病理学、眼科以及皮肤科等领域中表现出良好的性能<sup>[5-10]</sup>。Owais等<sup>[11]</sup>建立胃肠道疾病自动诊断系统,但其系统仅能分辨食管、胃、小肠、大肠和直肠,具体病变部位仍需依靠人工标注。因此,本研究希望构建一个上消化道解剖分类模型,减轻人工标注的负担。

本研究收集苏州大学附属第一医院消化内镜中心的胃镜图片,利用在ImageNet数据库预训练的3种深度卷积神经网络Xception<sup>[12]</sup>、NASNet Large (NASNet)<sup>[13]</sup>和ResNet50V2 (ResNet)<sup>[14]</sup>,通过对胃镜图片进行迁移学习,构建上消化道内镜解剖分类模型,用以区分食管、贲门、胃底、胃体、胃角、胃窦、幽门、十二指肠球部及降部。

1 资料与方法

1.1 胃镜图片收集与解剖位置标注

本研究收集4 453张胃镜图片。其中,4 183张图片来自苏州大学附属第一医院消化内镜中心,包含食管589张、贲门370张、胃底726张、胃体615张、胃角329张、胃窦448张、幽门295张、十二指肠球部351张、十二指肠降部460张。其余270张胃镜图片来自江苏大学附属金坛医院,每个部位图片各30张。图片解剖位置标注根据细井董三《标准胃镜检查》<sup>[15]</sup>中

的采图要求结合胃的解剖,具体解剖位置标注如图1所示。

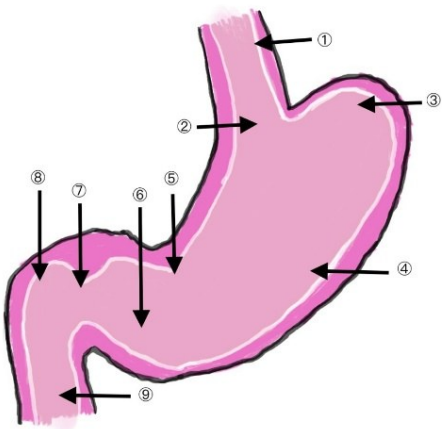


图1 上消化道解剖分类

Figure 1 Anatomical classification of the upper gastrointestinal tract  
①食管;②贲门;③胃底;④胃体;⑤胃角;⑥胃窦;⑦幽门;⑧十二指肠球部;⑨十二指肠降部

1.2 胃镜图片预处理

考虑到原始胃镜图片尺寸不一且像素较高,直接输入深度卷积神经网络训练易导致网络训练耗时过久与欠拟合。为提升深度卷积神经网络的训练效率,将纳入的4 453张内镜图片统一至331×331像素,以RGB三通道形式保存,其保存格式为[331, 331, 3]。

1.3 上消化道内镜解剖分类模型的构建

将经过预处理的来自苏州大学附属第一医院消化内镜中心的4 183张胃镜图片按8:2比例随机分为训练集和验证集,来自江苏大学附属金坛医院的270张图片作为外部测试集构建模型,如图2所示。上消化道内镜解剖分类模型的构建过程如下:(1)深度卷积神经网络架构选择:选取Xception、NASNet和ResNet 3种被广泛使用的架构,用于训练上消化道内镜解剖分类模型,3种架构均已于Imagenet数据库完成预训练,载入预训练权重;(2)架构修改:为适应本

研究对上消化道内镜解剖分类的任务目标,对网络架构采取适当修改,3种架构的原全连接层是基于Imagenet数据库1000分类所设置,不适用于本研究9个解剖位置分类,遂将原架构全连接层重塑,统一增加3层全连接层(每层节点数由上至下为2048、1024、512),模型输出层选用激活函数为softmax函数,以适应多分类任务;(3)模型训练:载入训练集数据于3种不同架构中,进行上消化道内镜解剖分类模型的训练,考虑到本研究样本量较少,直接训练深度卷积神经网络易导致欠拟合,故本研究的模型训练采用迁移学习中特征提取策略,即固定除全连接层外其余层的预训练权重,利用反向传播算法,重新训练全连接层权重;(4)模型训练超参数设定:3种卷积神经网络架构,训练轮次均为50,训练批次为32,学习率为0.001,优化函数为SGD算法(随机梯度下降算法)。

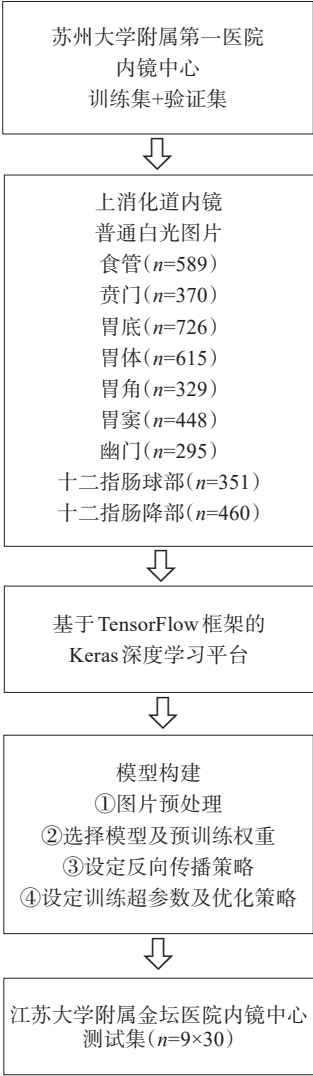


图2 上消化道内镜解剖分类模型构建的流程图

Figure 2 Flowchart of the development of anatomical classification models for upper gastrointestinal endoscopy

1.4 上消化道内镜解剖分类模型的评价

运用验证集及测试集数据评价上述获得的3个上消化道解剖部位图片分类模型。各个模型的分类结果采用以下评价指标:准确性(accuracy)、召回率(recall)、精确度(precision)。

深度卷积神经网络在很多任务上表现出了较高的准确性,但是其学习到的内容及其规则却难以解释,因此,本研究采用梯度加权分类激活映射(Grad-CAM)算法对模型分类结果进行可视化解释。步骤如下:(1)随机选取一张经预处理的待分类图片输入模型;(2)获取最后一卷积层激活输出的梯度;(3)表征出最后卷积层激活输出各点位对模型决策分类的重要程度;(4)通过ReLU函数剔除负梯度信息,得到类激活热力图;(5)将原始内镜图与热力图叠加从而对模型分类结果进行可视化分析。

2 结果

2.1 上消化道内镜解剖分类模型评价

分别在训练集、验证集和测试集中对构建好的3个上消化道内镜解剖分类模型进行模型分类能力评价。结果发现,各模型均具备较高的分类能力,其分类准确性与精确度都达到了较高水平。Xception模型在验证集中,平均分类准确性为0.980,平均分类召回率为0.895,平均分类精确度为0.921,AUC为0.991;在测试集中,平均分类准确性为0.985,平均分类召回率为0.926,平均分类精确度为0.936,AUC为0.998。NASNet模型在验证集中,平均分类准确性为0.977,平均分类召回率为0.883,平均分类精确度为0.906,AUC为0.987;在测试集中,平均分类准确性为0.992,平均分类召回率为0.959,平均分类精确度为0.970,AUC为0.996。ResNet模型在验证集中,平均分类准确性为0.982,平均分类召回率为0.905,平均分类精确度为0.933,AUC为0.991;在测试集中,平均分类准确性为0.987,平均分类召回率为0.941,平均分类精确度为0.944,AUC为0.998。

在验证集中,平均分类准确性为0.980,平均分类召回率为0.894,平均分类精确度为0.920;其中,ResNet模型表现最优,其分类准确性(0.982)、分类召回率(0.905)和分类精确度(0.933)最高。在测试集中,平均分类准确性为0.988,平均分类召回率为0.942,平均分类精确度为0.950;其中,NASNet模型表现最优,其分类准确性(0.992)、分类召回率(0.959)和分类精确度(0.970)最高。

2.2 上消化道内镜解剖分类模型分类结果的可视化解释

本研究采用Grad-CAM算法,将获得的类激活热



力图与原始胃镜图片叠加,从而实现对模型分类结果进行可视化分析。结果如图3所示,图中a列为内镜原图,b列为卷积神经网络特征提取层输出形成的热力矩阵图,c列为最终模型输出热力图覆盖于内镜原图后的效果图。热力图中红黄色高亮部分,代表着深度学习模型判断该图片属于胃镜解剖位置的依

据部分。例如图a5、b5、c5所示,卷积神经网络提取出的热力图中高亮部分重合于胃镜图片中的胃角位置,直观地表明了模型的分类依据。图a7、b7、c7可以看到模型判定该胃镜图片为幽门的依据是其张开的幽门口。由此可见,热力图的应用可以更直观地了解模型分类的依据,更好地评估模型分类能力。

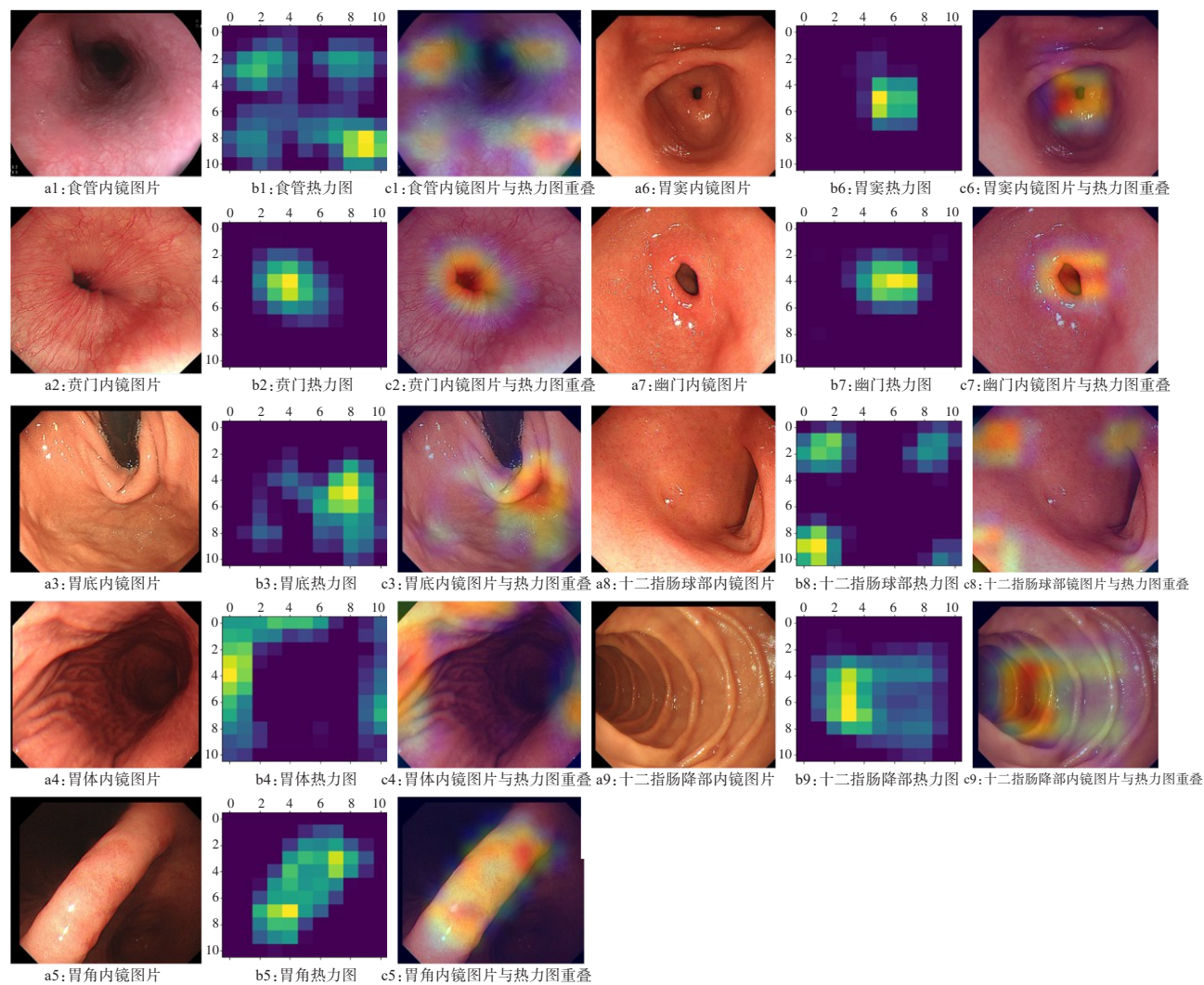


图3 上消化道内镜解剖分类模型的可视化解释

Figure 3 Visual interpretation of anatomical classification models for upper gastrointestinal endoscopy

本研究选取了部分模型分类错误的图片进行分析,了解其错误原因。结果如图4所示,a1标记为幽门,模型判读为食管。错误原因:光线差、胃腔内泡沫较多且幽门未打开。b1标记为胃窦,模型判读为幽门。错误原因:胃窦和幽门的标记本身就存在一定混淆性,另外有部分胃角也显示在此图片中,因此标记为胃窦。

3 讨论与结论

近年来,深度学习在内窥镜中应用广泛,研究表

明卷积神经网络在Barrett食管、食管癌、幽门螺杆菌胃炎、胃癌的辅助诊断中表现出良好的性能<sup>[16-19]</sup>。目前也有关于深度学习用于上消化道内镜解剖定位的研究。Takiyama等<sup>[20]</sup>将胃镜图像分为喉、食管、胃、十二指肠4个部分并进一步将胃分为上、中、下3类。Igarashi等<sup>[21]</sup>构建一个分类模型,收集了85 246张图片,将其分为4个部分(咽喉、食管、胃和十二指肠),并根据成像模式分为白光和窄带成像,最后根据是否存在着色物质、碘染色和其他可处理方式对图像进行分类,共将其分为14类。Cogan等<sup>[22]</sup>构建了一

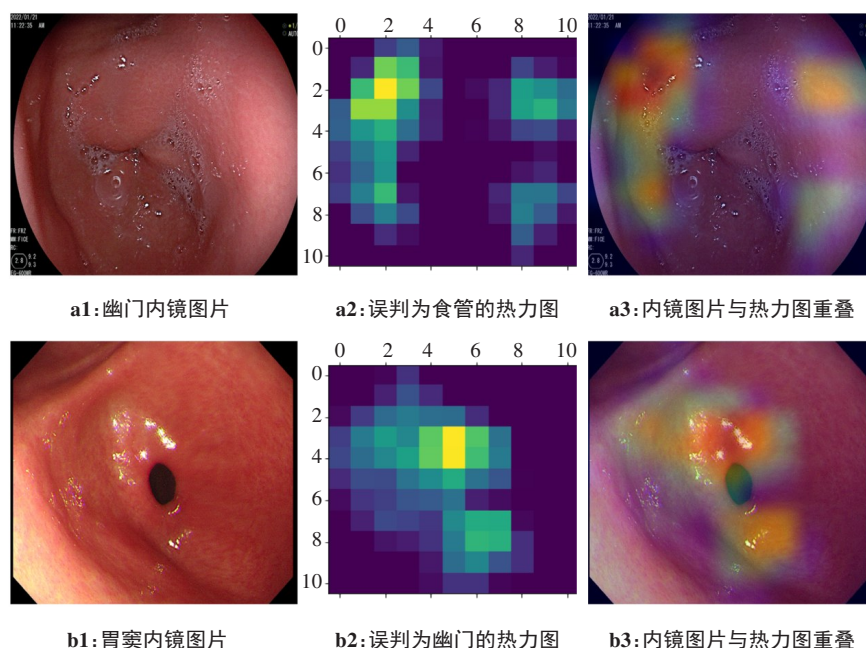


图4 上消化道内镜解剖分类模型错误分类示例

Figure 4 Incorrect classification examples of anatomical classification models for upper gastrointestinal endoscopy

个由8个类别组成的分类模型,其中3个代表解剖标志(幽门、齿状线、盲肠),3个代表病理状态(食管炎、溃疡性结肠炎、息肉),两个与病变切除有关。这些研究进一步证明了深度学习在上消化道内镜解剖定位中的潜力,但由于其解剖部位分类较为笼统,临床应用有限。本研究结合EGSE指南<sup>[23]</sup>和一些后续的指南与建议<sup>[24]</sup>,将解剖定位分类为食管、贲门、胃底、胃体、胃角、胃窦、幽门、十二指肠球部和十二指肠降部,更符合临床应用。

本研究是一项回顾性研究,收集苏州大学附属第一医院消化内镜中心图片,利用迁移学习,将已在ImageNet预训练过的Xception、NASNet和ResNet网络作为初始网络,构建了上消化道内镜解剖分类模型。在3个模型中,ResNet模型在验证集表现最优,拥有0.982的分类准确性、0.905的分类召回率和0.933的分类精确度;NASNet模型在测试集表现最优,拥有0.992的分类准确性、0.959的分类召回率和0.970的分类精确度。He等<sup>[25]</sup>在2020年构建的食管、贲门、胃底、胃体、胃角、胃窦、球部、降部八分类模型的分类准确性为94%。Chang等<sup>[26]</sup>在2021年构建了喉部、食管、食管胃交界、胃底、胃体、胃角、胃窦、十二指肠球部、十二指肠降部的九分类模型,其基础模型准确率为96.29%,增强模型准确率为96.64%。Choi等<sup>[27]</sup>构建了一个食管、贲门、胃底、胃体、胃角、胃窦、十二指肠球部、十二指肠降部的九分类CNN模型,达到了97.58%的准确率。本研究建立的模型最佳准确性在验证集上达到98.2%,高于这些相似解剖

分类的研究。本研究在收集图片时并未对一些存在阻碍分类可能特征的图像(如炎症、息肉、胃液、血液等)进行排除,因为这在临床实践中较为常见,以便由此训练出的模型在临床中有效。本研究还收集了江苏大学附属金坛医院的图片作为测试集对模型进行外部验证,得到了99.2%的准确性,表明本研究的分类模型具有良好的泛化能力。本研究采用梯度加权分类激活映射算法对模型分类结果进行可视化解释,以便于了解模型的工作原理。

本研究尚存在不足,本研究的样本量较小。虽然使用迁移学习,使得本研究的分类模型具有较高的准确性,但更大的样本量可提升分类模型性能。后期可以继续收集多中心数据,以期模型性能得到进一步提升。本研究仅收集了白光内镜下的图片,后续可以收集放大内镜、色素内镜下的图片对模型进行进一步训练,增加模型泛用性。

## 【参考文献】

- [1] Park WG, Shaheen NJ, Cohen J, et al. Quality indicators for EGD[J]. Am J Gastroenterol, 2015, 110(1): 60-71.
- [2] 李华, 高天华, 霍志斌, 等. 早期上消化道肿瘤的内镜诊断进展[J]. 中国研究型医院, 2022, 9(2): 56-60.  
Li H, Gao TH, Huo ZB, et al. Early of endoscopic diagnosis of upper gastrointestinal tumor progress[J]. China Research Hospital, 2022, 9(2): 56-60.
- [3] Asge Standards of Practice Committee, Faulx AL, Lightdale JR, et al. Guidelines for privileging, credentialing, and proctoring to perform GI endoscopy[J]. Gastrointest Endosc, 2017, 85(2): 273-281.
- [4] Asge Training Committee, Adler DG, Bakis G, et al. Principles of training in GI endoscopy[J]. Gastrointest Endosc, 2012, 75(2): 231-235.
- [5] Chan HP, Samala RK, Hadjiiski LM, et al. Deep learning in medical

- image analysis[J]. *Adv Exp Med Biol*, 2020, 1213: 3-21.
- [6] Samala RK, Chan HP, Hadjiiski LM, et al. Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms[J]. *Phys Med Biol*, 2017, 62(23): 8894-8908.
- [7] Masood A, Sheng B, Li P, et al. Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images[J]. *Phys Med Biol*, 2018, 79: 117-128.
- [8] Du Y, Zhang R, Zargari A, et al. Classification of tumor epithelium and stroma by exploiting image features learned by deep convolutional neural networks[J]. *Ann Biomed Eng*, 2018, 46(12): 1988-1999.
- [9] Liu X, Jiang J, Zhang K, et al. Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network [J]. *PLoS One*, 2017, 12(3): e0168606.
- [10] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. *Nature*, 2017, 542(7639): 115-118.
- [11] Owais M, Arsalan M, Mahmood T, et al. Automated diagnosis of various gastrointestinal lesions using a deep learning-based classification and retrieval framework with a large endoscopic database: model development and validation[J]. *J Med Internet Res*, 2020, 22(11): e18563.
- [12] Chollet F. Xception: deep learning with depthwise separable convolutions[J/OL]. [2021-12-28]. <http://arxiv.org/abs/1610.02357>.
- [13] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[J/OL]. [2021-12-28]. <http://arxiv.org/abs/1512.03385>.
- [14] Zoph B, Vasudevan V, Shlens J, et al. Learning transferable architectures for scalable image recognition[J/OL]. [2021-12-30]. <http://arxiv.org/abs/1707.07012v4>.
- [15] 细井董三. 标准胃镜检查[M]. 沈阳: 辽宁科学技术出版社, 2013. Hosoidaisho SA. Standard gastroscopy[M]. Shenyang: Liaoning Science and Technology Press, 2013.
- [16] De Groof AJ, Struyvenberg MR, Van Der Putten J, et al. Deep-learning system detects neoplasia in patients with Barrett's esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking[J]. *Gastroenterology*, 2020, 158(4): 915-929.
- [17] Kumagai Y, Takubo K, Kawada K, et al. Diagnosis using deep-learning artificial intelligence based on the endocytoscopic observation of the esophagus [J]. *Esophagus*, 2019, 16(2): 180-187.
- [18] Shichijo S, Nomura S, Aoyama K, et al. Application of convolutional neural networks in the diagnosis of helicobacter pylori infection based on endoscopic images[J]. *eBioMedicine*, 2017, 25: 106-111.
- [19] Hirasawa T, Aoyama K, Tanimoto T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images[J]. *Gastric Cancer*, 2018, 21(4): 653-660.
- [20] Takiyama H, Ozawa T, Ishihara S, et al. Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks[J]. *Sci Rep*, 2018, 8(1): 7497.
- [21] Igarashi S, Sasaki Y, Mikami T, et al. Anatomical classification of upper gastrointestinal organs under various image capture conditions using AlexNet[J]. *Comput Biol Med*, 2020, 124: 103950.
- [22] Cogan T, Cogan M, Tamil L. MAPGI: accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning[J]. *Comput Biol Med*, 2019, 111: 103351.
- [23] Rey JF, Lambert R, Esge Quality Assurance Committee. ESGE recommendations for quality control in gastrointestinal endoscopy: guidelines for image documentation in upper and lower GI endoscopy [J]. *Endoscopy*, 2001, 33(10): 901-903.
- [24] Marques S, Bispo M, Pimentel-Nunes P, et al. Image documentation in gastrointestinal endoscopy: review of recommendations[J]. *GE Port J Gastroenterol*, 2017, 24(6): 269-274.
- [25] He Q, Bano S, Ahmad OF, et al. Deep learning-based anatomical site classification for upper gastrointestinal endoscopy[J]. *Int J Comput Assist Radiol Surg*, 2020, 15(7): 1085-1094.
- [26] Chang YY, Li PC, Chang RF, et al. Deep learning-based endoscopic anatomy classification: an accelerated approach for data preparation and model validation[J]. *Surg Endosc*, 2022, 36(6): 3811-3821.
- [27] Choi SJ, Khan MA, Choi HS, et al. Development of artificial intelligence system for quality control of photo documentation in esophagogastroduodenoscopy[J]. *Surg Endosc*, 2022, 36(1): 57-65.

(编辑:陈丽霞)