

## 机器学习在食管癌诊疗中的研究进展

王泳荃, 苏梦琦, 石清磊, 马艺宁, 孙扬帆, 王昌淼, 汪国有, 裘肖明, 尹义龙, 万翔

### 引用本文

王泳荃, 苏梦琦, 石清磊, 马艺宁, 孙扬帆, 王昌淼, 汪国有, 裘肖明, 尹义龙, 万翔. [机器学习在食管癌诊疗中的研究进展](#)[J]. 计算机科学, 2025, 52(9): 4-15.

WANG Yongquan, SU Mengqi, SHI Qinglei, MA Yining, SUN Yangfan, WANG Changmiao, WANG Guoyou, XI Xiaoming, YIN Yilong, WAN Xiang. [Research Progress of Machine Learning in Diagnosis and Treatment of Esophageal Cancer](#) [J]. Computer Science, 2025, 52(9): 4-15.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

### Similar articles recommended (Please use Firefox or IE to view the article)

#### [数据分类分级技术研究综述](#)

Survey of Data Classification and Grading Studies

计算机科学, 2025, 52(9): 195-211. <https://doi.org/10.11896/jsjcx.240800149>

#### [基于动态平衡和距离抑制的点云语义分割主动学习](#)

Active Learning for Point Cloud Semantic Segmentation Based on Dynamic Balance and Distance Suppression

计算机科学, 2025, 52(8): 180-187. <https://doi.org/10.11896/jsjcx.240900104>

#### [基于改进SOM网络的聚类算法](#)

Clustering Algorithm Based on Improved SOM Model

计算机科学, 2025, 52(8): 162-170. <https://doi.org/10.11896/jsjcx.240700017>

#### [基于深度学习的混凝土缺陷检测方法综述](#)

Review of Concrete Defect Detection Methods Based on Deep Learning

计算机科学, 2025, 52(6A): 240900137-12. <https://doi.org/10.11896/jsjcx.240900137>

#### [森林火灾风险预测的研究进展及面临的挑战](#)

Research Progress and Challenges in Forest Fire Risk Prediction

计算机科学, 2025, 52(6A): 240400177-8. <https://doi.org/10.11896/jsjcx.240400177>

# 机器学习在食管癌诊疗中的研究进展

王泳荃<sup>1</sup> 苏梦琦<sup>2</sup> 石清磊<sup>3,4</sup> 马艺宁<sup>5</sup> 孙扬帆<sup>5</sup> 王昌淼<sup>4</sup> 汪国有<sup>1</sup> 裘肖明<sup>6</sup> 尹义龙<sup>3</sup>  
万翔<sup>4</sup>

1 华中科技大学人工智能与自动化学院 武汉 430074

2 新加坡国立大学计算机学院 新加坡 119077

3 山东大学软件学院 济南 250101

4 深圳市大数据研究院 广东 深圳 518172

5 香港中文大学(深圳)理工学院 广东 深圳 518172

6 山东建筑大学计算机科学与技术学院 济南 250101

(wangyongquan@hust.edu.cn)

**摘要** 食管癌(Esophageal Cancer, EC)是一种全球范围内高致死率的恶性肿瘤,尤其是在我国,由于早期诊断率低、预后不良,食管癌已成为临床诊疗中面临的重大挑战。近年来,机器学习(Machine Learning, ML)技术凭借多模态数据融合的智能分析方法,在推动食管癌诊疗精准化发展方面取得了显著进展。传统机器学习方法通过整合食管癌影像学特征与临床文本信息,有效提升了早期病变诊断的敏感性,并为高风险患者的分层管理提供了科学支持。卷积神经网络(Convolutional Neural Network, CNN)以其高效的参数共享机制和卓越的局部特征提取能力,进一步增强了食管癌早期诊断与筛查的准确性。此外,将CNN与基于自注意力机制的Transformer模型相结合,显著提升了全局特征的建模能力,通过多模态数据的协同作用,其在食管癌病灶分割、早期诊断、疗效预测和生存分析等方面展现出广阔的应用前景。然而,食管癌病变的高度异质性以及图像数据类别不平衡问题,依然对机器学习技术的临床应用带来了较大挑战。为进一步推动食管癌智能诊疗技术的发展,聚焦于食管癌早期筛查与诊断、疗效预测与生存分析、影像分割3个关键领域,系统综述了传统机器学习、CNN及Transformer等前沿技术在EC诊疗中的研究现状与挑战,旨在为未来食管癌智能化诊疗研究提供有价值的参考与借鉴。

**关键词:** 食管癌;机器学习;鉴别诊断;生存分析;语义分割

中图分类号 TP391

## Research Progress of Machine Learning in Diagnosis and Treatment of Esophageal Cancer

WANG Yongquan<sup>1</sup>, SU Mengqi<sup>2</sup>, SHI Qinglei<sup>3,4</sup>, MA Yining<sup>5</sup>, SUN Yangfan<sup>5</sup>, WANG Changmiao<sup>4</sup>, WANG Guoyou<sup>1</sup>, XI Xiaoming<sup>6</sup>, YIN Yilong<sup>3</sup> and WAN Xiang<sup>4</sup>

1 College of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China

2 College of Computing, National University of Singapore, Singapore 119077, Singapore

3 School of Software, Shandong University, Jinan 250101, China

4 Shenzhen Research Institute of Big Data, Shenzhen, Guangdong 518172, China

5 School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Shenzhen, Guangdong 518172, China

6 School of Computer Science and Technology, Shandong Jianzhu University, Jinan 250101, China

**Abstract** EC is a highly lethal malignancy worldwide, particularly in China. Its low early diagnosis rates and poor prognosis present significant challenges to clinical management. In recent years, ML technology, based on multi-modal data, has shown great potential in early EC diagnosis and treatment. Traditional ML methods integrate radiomic features from EC imaging and clinical textual data, thereby effectively improving the sensitivity of early lesion diagnosis and providing scientific support for the stratified management of high-risk patients. CNNs with their efficient parameter-sharing mechanisms and excellent local feature extraction capabilities, further enhance the accuracy of early EC diagnosis and screening. Moreover, combining CNNs with Transformer models based on self-attention mechanisms significantly strengthens global feature modeling, which demonstrating broad applica-

到稿日期:2025-01-10 返修日期:2025-02-20

基金项目:2023 大学发展基金-科研启动经费即食管癌的图像质量控制和智能诊断系统;福田健康医疗研究项目(FTWS002)

This work was supported by the 2023 University Development Fund-Scientific Research Start-up Grant Image Quality Control and Intelligent Diagnostic System for Esophageal Cancer and Futian Health and Medical Research Project(FTWS002).

通信作者:石清磊(shiqingleicmu@gmail.com)

tion potential in lesion segmentation, early diagnosis, treatment outcome prediction, and survival analysis of EC. However, the high heterogeneity of EC lesions and the class imbalance of imaging data continue to pose significant challenges to the clinical application of ML technologies. To advance intelligent diagnostic and therapeutic technologies for EC, this paper focuses on three critical areas: early screening and diagnosis, treatment outcome prediction and survival analysis, and image segmentation. It systematically reviews the current research status and challenges of traditional ML, CNNs, and emerging Transformer technologies in EC diagnosis and treatment, aiming to provide valuable insights and references for future research on intelligent EC diagnosis and treatment.

**Keywords** Esophageal cancer, Machine learning, Differential diagnosis, Survival analysis, Semantic segmentation

## 1 引言

根据世界卫生组织(World Health Organization, WHO)在2019年的评估,癌症已成为绝大多数国家排名前两位的致死因素。据2020年全球癌症统计报告显示,2040年全球的新发癌症病例数将比2020年高出47%,达2840万例,于我国目前的经济发展水平而言,癌症在未来几年将带来较大负担。其中食管癌(Esophageal Cancer, EC)在全球范围内的发病率和死亡率分别排名第七和第六<sup>[1]</sup>,而在中国的发病率和死亡率分别排名第六和第四,均高于全球平均水平<sup>[2]</sup>,如图1所示。随着人口老龄化加剧,我国EC比率进一步上升,因此,提高食管癌的诊断准确率和远期治疗效果,是亟待研究的重要课题<sup>[3]</sup>。

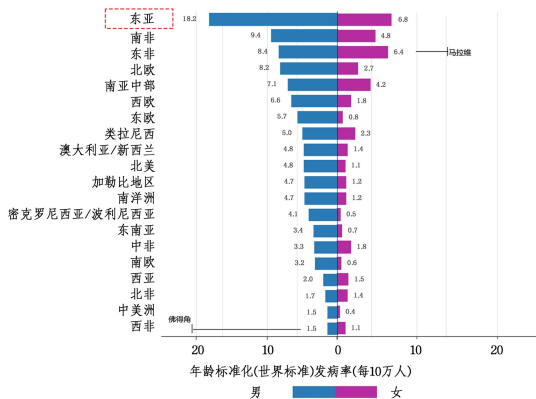


图1 食管癌发病率的区域差异<sup>[1]</sup>

Fig. 1 Regional differences in esophageal cancer incidence<sup>[1]</sup>

不同类型EC的治疗方式不同<sup>[4]</sup>,对于黏膜癌,内镜黏膜切除术和内镜黏膜下剥离术是标准疗法;而对于局部晚期癌症,食管切除术仍是主要方法;当肿瘤更为严重时,通常会加入新辅助化疗或新辅助放疗。将化疗加入EC治疗的优势在于,术前可能缩小肿瘤,同时靶向减小转移灶,从而降低远处转移风险。放疗是不可切除EC的标准治疗方法,并且也可作为可切除EC的治疗选择。对于因健康或技术原因无法手术的患者,同时放、化疗应作为标准护理。为提高EC的诊疗效果,针对每一个患者的特殊诊断结果,制定精准和个性化的治疗方案是临床迫切需要的。以往由于不同医生之间诊疗经验的差异,以及医疗资源的有限性,限制了治疗方案的制定水平和精准性。近年来,基于机器学习(Machine Learning, ML)的智能化诊疗技术在EC的临床诊疗中展现出了较大的应用价值。

在EC诊疗中,机器学习技术的应用呈现多层次的技术

演进与互补。传统机器学习方法凭借其在小样本结构化数据中的高效性,在早期筛查和疗效预测中展现了独特价值。然而,此类方法高度依赖人工特征工程,限制了其在复杂非结构化数据中的应用潜力,且难以捕捉数据中的深层关联,尤其在分割任务中表现受限。卷积神经网络(Convolutional Neural Network, CNN)通过自动提取局部特征的能力,显著提升了EC影像分析的精度与效率,在病灶分割中的表现尤为突出<sup>[5]</sup>。此外, CNN在多模态生存预测中的创新应用为个性化治疗提供了新思路。然而, CNN在建模全局上下文的能力不足,且“黑箱”特性导致其可解释性欠缺,制约了其在临床决策中的信任度。CNN与Transformer的混合模型通过融合局部特征与全局依赖关系,成为当前研究的前沿方向,此类模型在病理切片分类、多模态生存预测及复杂分割任务中展现出协同优势<sup>[6]</sup>。尽管如此,混合模型仍面临计算复杂度高、多模态交互机制不完善等挑战,且在多任务协同方面的能力尚未被充分挖掘,亟需轻量化设计与跨模态深度融合技术的突破。

总体而言,传统方法、CNN及混合模型在EC诊疗中形成了“简单任务高效性—复杂任务深层建模—全局上下文协同”的技术谱系。为了进一步推动EC智能化诊疗技术的进步,本文从EC早期筛查与诊断、疗效预测与生存分析,以及EC影像分割3个主要领域出发,综述了传统机器学习方法、卷积神经网络及Transformer等新兴技术在EC诊疗中的应用现状和面临的挑战,旨在为EC智能化诊疗的研究提供有益的参考。

## 2 机器学习

机器学习近年来在EC诊疗领域展现出了颠覆性的应用价值,其核心在于模拟和扩展人类智能。按照ML技术在EC中应用的技术类型,可将其划分为传统机器学习技术、卷积神经网络(CNN)技术以及以Transformer为代表的新兴技术。

传统机器学习技术是指在深度学习兴起之前,依赖手工特征工程和统计方法从数据中学习规律的技术。传统机器学习主要包括监督学习、无监督学习和半监督学习三大类别,在食管癌诊疗中主要用于结构化数据分析与临床决策支持。在监督学习中,模型通过已标注数据来学习输入与输出之间的映射关系,已被广泛应用于食管癌早期筛查场景,通过整合患者血清标志物与内镜窄带成像(NBI)中的血管形态特征(IPCL分型),可实现早期病变的自动化识别。无监督学习则不依赖标签数据,目标是从数据中发现隐藏结构或模式,在食管癌的诊疗中,该方法常用于对数据进行降维,以解决小样本

数据的过拟合问题。半监督学习结合了监督与无监督学习,利用少量标注数据和大量未标注数据共同训练模型。总体来说,传统机器学习技术仍在食管癌的诊疗中发挥着重要作用,尤其是在数据量较小或特征结构明确的情况下。

卷积神经网络(CNN)是深度学习领域的一种前馈神经网络,能够通过卷积操作从数据中自动提取特征,革新了食管癌影像分析的范式。CNN技术凭借其局部连接、权重共享和多层结构设计,通过局部感受野和权值共享机制来减少参数

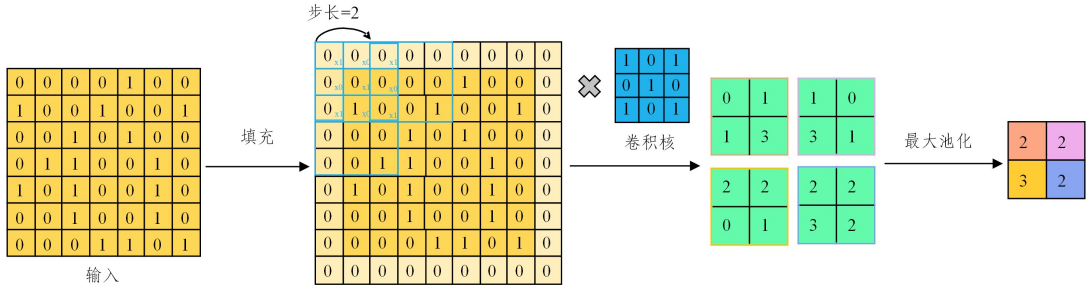


图2 二维卷积神经网络流程示意图<sup>[7]</sup>

Fig. 2 Schematic diagram of workflow for 2D CNN<sup>[7]</sup>

Transformer技术最初用于自然语言处理,凭借其独特的自注意力机制和并行计算能力,实现了对长距离依赖关系的建模,显著提升了机器学习在食管癌诊疗中的效果<sup>[8]</sup>。Transformer通过全局注意力机制捕捉数据中的长程依赖特征,相较于CNN,其不仅能提取局部信息,还能建模全局上下文,其基本结构如图3所示。近年来,Transformer与CNN的结合表现出较大潜力,CNN擅长局部特征提取,而Transformer则在全局信息整合和长程依赖建模中表现出优异性。二者的结合在捕获局部和全局信息、整合多模态数据以及提升模型性能方面表现卓越。在EC诊疗中,这种结合不仅提升了分类、分割和生存预测的精度,也为构建多任务诊疗模型提供了新的方向。

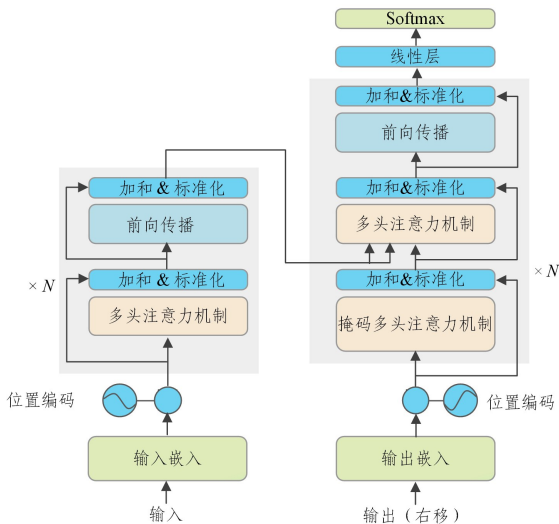


图3 Transformer模型架构<sup>[9]</sup>

Fig. 3 Architecture of Transformer model<sup>[9]</sup>

Transformer与CNN的结合虽然在食管癌的诊疗中表现出较大潜力,但是较高的计算复杂度也限制了其在临床中的应用效果,尤其在医疗数据量较小时。线性注意力机制

数量,提取数据的局部特征并逐层构建高层抽象表示,从而降低了过拟合风险,特别适合大规模非结构化数据的处理(如图像和视频)。典型的卷积神经网络包括以下要点:填充扩展输入,卷积核提取特征,步长控制移动幅度,最大池化降维,最终生成较小特征图保留关键信息。其流程如图2所示。近年来,CNN在EC诊疗中展现出广泛的应用价值,通过特征提取、注意力机制及多模态融合等方法,CNN显著提升了EC的诊断、治疗预测,以及EC分割的效率和精度。

(Linear Transformer, Linformer)是一种改进的自注意力计算方法,旨在突破传统Transformer模型中自注意力计算的效率瓶颈。传统的自注意力机制让每个token在计算时与其他token进行交互,使得在处理较长序列时,训练和部署Transformer模型需要大量的计算资源。Linformer的核心创新在于将自注意力矩阵近似为低秩矩阵,如图4所示。具体而言,它在应用缩放点积注意力前,先将Key和Value矩阵映射到较低维度的空间,从而将计算和空间复杂度从 $O(n^2)$ 降低到 $O(n)$ ,其中 $n$ 是序列长度。这一方法显著提高了效率,同时在性能上与标准Transformer模型相当<sup>[10]</sup>。实验表明,在医学图像任务(如食管癌生存预测)中,Linformer的精度损失可忽略<sup>[5]</sup>。与标准自注意力相比,Linformer能显著降低内存占用与计算成本,适用于长序列或多模态数据(如内镜视频时序分析),但也存在局限性,如低秩投影可能会损失高频细节,对病灶微结构敏感的任务需谨慎使用。

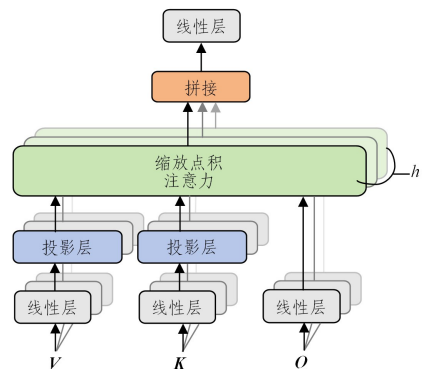


图4 多头线性自注意力机制架构<sup>[10]</sup>

Fig. 4 Architecture of multi-head linear self-attention<sup>[10]</sup>

### 3 机器学习在EC诊疗中的应用现状

机器学习技术按照其在EC诊疗中应用的技术类型可以

分为传统 AI 技术、CNN<sup>[7]</sup> 和基于 Transformer 的神经网络<sup>[11]</sup> 3 类,按照临床应用的场景和目的的不同,可以分为早期筛查和诊断、疗效预测和生存分析、EC 的分割 3 个方面。接下来,本文从传统 AI 技术、CNN 和基于 Transformer 的神

网络 3 种技术类型着手,分别综述它们在早期筛查和诊断、疗效预测和生存分析,以及 EC 的分割 3 种临床场景中的应用现状和挑战。图 5 展示了整个机器学习在 EC 诊疗中的发展阶段。

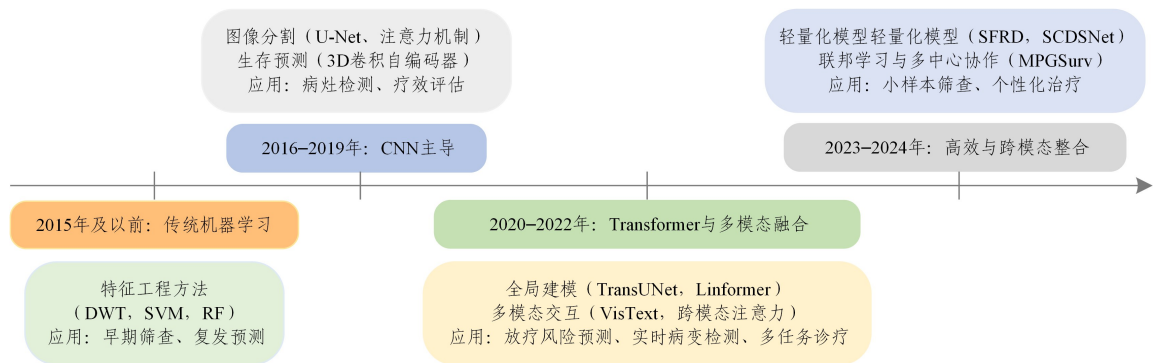


图 5 机器学习在 EC 诊疗中的技术发展

Fig. 5 Technological development of machine learning in EC diagnosis and treatment

### 3.1 传统机器学习技术在 EC 诊疗中的应用

传统机器学习算法是指深度学习兴起前广泛应用的算法,其依赖逻辑推理、专家系统、浅层模型结构及显式特征提取。该类算法通常在小数据集上表现良好,具有高解释性和计算效率,主要包括监督学习、无监督学习和强化学习 3 类。常见的分类器包括决策树(Decision Tree, DT)、支持向量机(Support Vector Machine, SVM)、K 近邻(K-Nearest Neighbor, KNN)、随机森林(Random Forest, RF)、朴素贝叶斯(Naive Bayesian, NB)和逻辑回归(Logistic Regression, LR)等<sup>[12]</sup>。传统机器学习技术在 EC 诊疗中展现出重要的应用价值。

首先,传统机器学习技术在 EC 早期筛查和诊断中展现显著价值。例如,Yamaguchi 等<sup>[13]</sup>提出了一种自动检测内窥镜图像中早期食管癌的方法,其将图像分解为 RGB 和亮度分量,并进行离散小波变换(Discrete Wavelet Transform, DWT),结合盒计数法(Box-Counting Method, BCM)计算分形维数(Fractal Dimension, FD)来检测异常区域,实验表明,该方法有效。Xue 等<sup>[14]</sup>提出了基于多层感知器(Multilayer Perceptron, MLP),并结合改进的经验小波变换(Improved Empirical Wavelet Transform, IEWT)和一致性同源性的方法,用于早期 EC 的内镜诊断,结果表明,该方法具有较高的应用价值。针对传统模型计算资源消耗较大的问题,Padha 等<sup>[15]</sup>提出了混合量子-经典神经网络(Hybrid Quantum-Classical Neural Network, HQCNN),该方法结合 MLP 和量子框架,通过优化电路参数来最小化误差,结果显示其在 EC 分类中具有较高准确性,展示了量子技术在医疗领域的应用潜力。

其次,传统机器学习技术在 EC 疗效预测和生存分析中也展现出重要的应用价值。例如,Lyu 等<sup>[16]</sup>开发了一种预测 EC 切除术后复发的机器学习模型,通过顺序前向选择过程构建不同预测窗口的风险模型,并结合逻辑回归(LR)、支持向量机(SVM)、随机森林(RF)、决策树(DT)和朴素贝叶斯(NB)5 种算法,增强了模型的稳健性。Su 等<sup>[17]</sup>提出了一种基于小样本的堆叠集成学习方法,用于预测免疫治疗患者的

生存情况,该方法整合了 5 个子学习器和 1 个元学习器,结果显示其在小样本随访数据条件下,具备有效指导临床决策的潜力。Sun 等<sup>[18]</sup>提出了一种基于自组织映射(Self-organizing Map, SOM)和支持向量机(SVM)集成的生存风险预测方法,通过使用径向基核函数(Radial Basis Function, RBF)和人工蜂群优化算法(Artificial Bee Colony, ABC),显著提高了预测准确性,提供了一种高效的生存风险评估工具。

尽管尚未发现机器学习在 EC 分割中的独立应用,但在多任务诊疗中其展现了显著价值。例如,Prince 等<sup>[19]</sup>提出了结合人工神经网络和 XGBoost 分类器的混合深度学习模型(Hybrid Deep Learning Model, HDLM),用于提高食管肿瘤分类的预测精度。通过引入互注意力模块,HDLM 在肿瘤分割任务中的表现也得到提升,识别率达 97.81%,Dice 系数达 87.91%,表明其在实际应用中具备较高的有效性。

传统机器学习技术在单模态数据场景中展现了基础应用价值,但无法处理非结构化影像数据。随着多模态数据融合技术的发展,传统方法通过特征拼接或决策层集成扩展了应用能力。例如,Ngiam 等<sup>[20]</sup>提出了多模态深度自编码器,联合学习影像与临床文本特征,但需严格数据对齐。Wu 等<sup>[6]</sup>通过逻辑回归融合 CT 影像与临床文本特征,将生存预测的 C-index 提升至 0.78(单模态 0.68)。Guan 等<sup>[21]</sup>设计的 Vis-Text 模块关联 CT 图像与临床文本,在放疗后食管癌预测中,AUC 达 0.85(单模态 0.76)。然而,传统方法在多模态场景中仍面临模态异构性对齐困难与计算效率瓶颈。

传统 ML 技术也存在明显的局限性:1)过度依赖特征工程,需人为设计特征,限制了其对复杂任务的处理能力;2)扩展性差,在大规模数据集上性能不如深度学习,尤其在数据规模和维度增大时,计算效率显著下降;3)处理非结构化数据(如图像、语音)能力有限,表现通常逊色于深度学习,尤其在多模态场景中,仍面临模态异构性对齐困难与计算效率瓶颈。因此,传统机器学习适用于简单任务和小规模数据集,在复杂、大规模数据上表现受限。为了更清晰地展示传统机器学习在 EC 诊疗中的情况,进行了梳理和汇总,如表 1 所列。

表1 传统的机器学习技术在 EC 中的应用

Table 1 Application of traditional ML technology in EC

网络类型	数据来源	数据类型	临床任务	缺陷和不足
离散小波变换 (DWT), 盒计数法 (BCM), 基于分形维数 (FD) 检测异常区域 <sup>[13]</sup>	私有数据集: Department of Internal Medicine, Saga University, Japan, 1 280 * 1 024 和 1 408 * 1 056 像素	内窥镜图像 (2D)	早期食管癌的自动检测	依赖人工设计特征, 对噪声敏感; 分形维数计算复杂度高, 对早期病变的微小特征捕捉能力有限, 误检率高
改进经验小波变换 (IEWT), 基于深度学习附在经验小波变换 (DL-CEWT) <sup>[14]</sup>	私有数据集: 中国日照市人民医院	内窥镜图像 (2D)	食管癌的早期检测	IEWT 对光照变化和图像伪影敏感; DL-CEWT 需大量标注数据, 在小样本场景下易过拟合; 特征提取过程缺乏可解释性
混合量子-经典神经网络 (Hybrid Quantum-Classical Neural Network), 量子卷积神经网络 (QCNN), 量子循环神经网络 (QRNN), 量子长短期记忆 (QLSTM) <sup>[15]</sup>	公共数据集: SWELL-KW 多模态数据集, Wesad 多模态数据集, psykose 数据集	时间序列数据	心理健康状态预测	硬件资源受限, 实际部署困难; 混合框架训练过程复杂, 收敛速度慢, 泛化性不足
传统机器学习模型 (包括逻辑回归、支持向量机、随机森林、决策树和朴素贝叶斯) <sup>[16]</sup>	私有数据集: 罗切斯特大学医学中心胸外科和前肠外科部门, 2009 至 2018 年间接受食管切除术的 260 名连续患者的回顾性研究	无图像: 患者特征图表	预测食管切除术后食管癌复发	依赖人工特征选择, 无法自动提取深层特征; 对非结构化数据 (如影像) 处理能力差; 模型在类别不平衡数据 (复发/非复发) 中表现不稳定
堆叠集成学习方法 (包含 5 个子学习器和 1 个元学习器) <sup>[17]</sup>	私有数据集: 厦门大学第一附属医院 (2017 年 1 月至 2021 年 5 月) 的 113 名接受免疫疗法的食管癌患者数据	无图像: 该数据集捕获了全面的患者信息	预测接受免疫疗法的食管癌患者的生存情况	复杂度高, 计算资源需求较大; 小样本数据下子学习器多样性不足, 集成效果有限; 缺乏对时间序列动态变化建模能力
自组织映射 (SOM) 聚类和支持向量机 (SVM) 集成, 遗传算法 (GA)、粒子群优化 (PSO) 和人工蜂群 (ABC) 算法优化 SVM 参数 <sup>[18]</sup>	私有数据集: 郑州轻工业大学电气与信息工程学院, 及第一附属医院食管癌预防与治疗国家重点实验室, 180 例患者信息的 17 项血液指标	无图像: 17 个血液指标	预测食管癌生存风险水平	SOM 聚类对初始参数敏感, 稳定性差; SVM 在高维数据中核函数选择困难; 优化算法收敛速度慢, 适应实时临床需求难
混合深度学习模型 (HDLM), 包含人工神经网络 (ANN) 和 XGBoost 分类器 <sup>[19]</sup>	私有数据集: 上消化道内窥镜图像, 包含正常、食管炎和食管癌的图像, 共 1 500 张	内窥镜图像 (2D)	食管病变的多分类和分割	模型融合导致推理延迟增加; XGBoost 对超参数调优依赖性强; 多任务训练中分割与分类任务可能相互干扰, 需精细设计损失权重

### 3.2 CNN 技术在 EC 诊疗中的应用

CNN 是一种前馈神经网络, 能够通过卷积操作从数据中自动提取特征<sup>[7]</sup>。卷积操作是 CNN 的核心步骤, 其优点在于可以通过自动学习特征来处理复杂数据, 其采用局部连接

和权重共享机制显著减少了参数量, 降低了过拟合的风险。近年来, CNN 在食管癌的诊疗中表现出较大的应用价值。下面对 CNN 在 EC 的早期筛查和诊断、EC 疗效预测和生存分析、EC 的分割 3 个方面的应用现状进行介绍, 如表 2 所列。

表2 CNN 技术在 EC 中的应用

Table 2 Application of CNN technology in EC

网络类型	数据来源	数据类型	临床任务	缺陷和不足
基于深度学习 (DL) 技术的 DD-BA 方法, 包含 HRDEN (深度估计网络) 4 通道 Deeplabv3+ (语义分割网络) <sup>[22]</sup>	私有数据集: 西华大学附属医院消化内镜中心, 共 4 231 张食管图像, 来自 732 名患者	胃镜图像 (2D)	早期食管癌 (EEC) 病变的自动标注	深度图生成依赖预训练模型, 泛化性受限; 后处理步骤复杂, 难以端到端优化; 对小尺寸病灶 (小于 5 mm) 标注精度低
LIME 的可解释 AI (XAI) 技术 (CNN+迁移学习) <sup>[5]</sup>	公开数据集: Kvasir-V2 数据集, 包含 8 000 张图像, 分为 8 个不同类别	内窥镜图像 (2D)	食管恶性肿瘤的早期检测和分类	LIME 解释结果可能与实际决策逻辑不一致; 迁移学习模型参数量大, 计算成本高; 对低质量图像鲁棒性差
EPS YOLO-V5 (基于极化正则化和网络剪枝的端到端食管癌病变检测网络) <sup>[23]</sup>	私有数据集: 四川大学华西医院提供的内镜视频, 共 1 130 张食管癌窄带成像内窥镜图像	内镜视频逐帧提取的窄带成像内窥镜图像 (2D)	食管癌病变检测	剪枝后模型对罕见病变类型检测能力下降; 极化正则化增加训练时间; GhostNet 模块特征表达能力弱于原 YOLO-V5
自监督特征表示蒸馏 (Self-supervised Feature Representation Distillation, SFRD) <sup>[24]</sup>	私有数据集: 包含 2 000 张带标注的内镜食管图像; 额外 6 882 张无标签的内镜食管图像	窄带成像 (NBI) 食管内窥镜图像 (2D)	食管癌筛查	自监督预训练需大量无标签数据, 实际场景中获取困难; 教师-学生模型知识蒸馏可能引入噪声; 对黏膜微血管变化的敏感性不足
Transfer Learning with CNNs (将 SVM, RF 分类器与 4 个最广泛使用的网络 AlexNet, VGG16, VGG19 和 GoogLeNet 结合) <sup>[25]</sup>	公开数据集: Early Barrett's Cancer Detection Challenge 数据集 (MIC-CAI 2015 EndoVis Challenge 的子挑战) 100 张高清内窥镜图像	内窥镜图像 (2D)	早期食管腺癌 (EAC) 的自动检测和标注	预训练模型 (如 VGG) 对医学图像域适应性差; 浅层特征迁移导致细节信息丢失; 分类器 (如 SVM) 无法端到端优化
增强卷积神经网络 ECNN (CNN+SVM) <sup>[26]</sup>	公开数据集: KVASIR 数据集	内窥镜图像 (2D)	食管癌早期检测和分类	SVM 分类器在高维特征空间中表现不稳定; ECNN 的注意力模块增加过拟合风险; 对黏膜炎症等非癌变区域的误判率高
YOLO-V5 <sup>[27]</sup>	私有数据集: 7 例鳞状细胞癌 (SCC); 9 例异型增生 (Dysplasia); 10 例正常食管	白光图像 (WLI) 和窄带图 (NBD) (2D)	食管癌早期识别	YOLO-V5 对微小病变 (小于 3 mm) 漏检率高; 模型在动态内镜视频中帧间一致性差; 需大量标注数据支持, 标注成本较高
基于语义级别多模态数据融合的癌症筛查网络 (Semantic-level Cancer-Screening Network, SCNET) 框架 <sup>[28]</sup>	私有数据集: 来自安徽医科大学第一附属医院 (2016 年 7 月至 2017 年 9 月), 包括上消化道内窥镜图像和相应的文本报告	上消化道内窥镜图像和文本报告数据 (2D)	上消化道癌症筛查	文本与图像特征对齐困难, 语义鸿沟显著; 模型对文本报告的语言依赖性高, 实际场景中非标准化文本处理能力差; 多模态融合模块复杂度高, 推理速度慢

(续表)

网络类型	数据来源	数据类型	临床任务	缺陷和不足
3Dc 时序 DenseConvLstm Faster R-CNN, 结合了 3D 卷积神经网络(3DCNN)和卷积长短时记忆网络(ConvLstm) <sup>[29]</sup>	私有数据集:食管内窥镜视频数据集,44 个内窥镜视频,42 425 帧,包括正常、癌前病变(BE)和癌症(EAC 和 SCC)样本公开数据集: CVC-ClinicVideoDB 数据集(用于额外验证)	内窥镜视频数据(3D 视频数据由 2D 图像序列组成)	食管异常检测(包括癌前病变和癌症)	3D 卷积和 LSTM 模块计算成本较高,难以实时处理长视频;时序建模依赖帧间连续性,对动态模糊或镜头抖动敏感;罕见异常样本检测召回率低
基于深度学习的生存预测框架,包括 3D Coordinate Attention Convolutional Autoencoder (CACA) 和 Uncertainty-based Jointly Optimizing Cox Model(UOCM) <sup>[30]</sup>	私有数据集:来自福建医科大学附属泉州第一医院(2000 年至 2018 年),285 名食管癌患者的 CT 系列和临床信息	手动标注的 Gross Tumor Volume(GTV)区域的 CT 图像(3D)和临床信息(2D)	食管癌患者生存预测	3D 卷积自编码器参数量大,小样本数据下易过拟合;临床信息与影像特征融合策略简单,未建模跨模态交互;不确定性优化损失函数调参复杂,收敛稳定性差
多模态自注意力网络(编解码器+注意力机制) <sup>[21]</sup>	私有数据集:2014—2019 年间收集的食管癌患者数据,共 553 名患者,其中 186 名发展为食管癌	胸部计算机断层扫描 CT 图像(3D)和临床表格数据(2D)	预测食管癌患者放疗后发生食管癌的风险	自注意力模块对高维临床数据建模效率低;模态间交互缺乏可解释性;模型在多中心数据中泛化性受限于数据分布差异
基于生存网络的多模态人口图谱(MPGSurv) <sup>[31]</sup>	私有数据集:来自四川癌症医院,共 1 354 名 ESCC 患者的数据	原发肿瘤和淋巴结的 CT 影像(3D)临床血液学参数和随访信息(2D)	食管鳞状细胞癌(ESCC)患者的预后风险预测	图神经网络(GNN)需预定义模态关联拓扑,主观性强;血液学参数与影像特征的异质性融合不充分;模型训练依赖大规模数据,小样本场景性能下降显著
通道注意力 U-Net <sup>[32]</sup>	私有数据集:来自中山大学肿瘤防治中心,包含 152 名食管癌患者的胸部 CT 扫描,总共 13 050 张 CT 图像,被手动标注	2D CT 切片输入,最终进行 3D 重建(2D 和 3D 结合)	食管和食管癌的分割	通道注意力模块对低对比度肿瘤区域敏感性不足;跨层特征融合引入冗余计算;3D 重建依赖切片插值,空间连续性建模不精确
基于编码器-解码器架构的 2.5D 分割网络 Eso-Net <sup>[33]</sup>	私有数据集:来自中山大学肿瘤防治中心,30 名食管癌患者的 CT 扫描图像,共 6 362 张 CT 切片	3 张连续的 2D CT 切片输入,通道方式的 2.5D 分割(2.5D)	癌变食管的自动分割	2.5D 输入(3 张连续切片)导致局部上下文信息丢失;空洞卷积增大感受野的同时引入网格伪影;先验注意力模块依赖人工标注关键区域,自动化程度低
基于双编码器结构的 ARTSEG(Prior Anatomy and RT Information Enhanced Second-course Esophageal GTV Segmentation Network),包含区域保留注意力模块(Region-Preserving Attention Module,RAM) <sup>[34]</sup>	私有数据集:配对的第一、第二疗程数据集(Sp):来自中山大学肿瘤防治中心;GTV 数据集(Sv):来自美迈科技有限公司公开数据集,食管分割数据集(Se):来自 SegTHOR 公开数据集	CT 图像(3D)	食管肿瘤分割(GTV)	区域保留注意力对肿瘤形态变化适应性差;多中心数据分布差异导致模型校准困难;二次放疗中解剖结构变形影响分割精度
扩张密集注意力 U-Net(Dilated Dense Attention Unet,DDAUnet),基于卷积神经网络(CNN)的全自动端到端食管癌分割方法 <sup>[35]</sup>	私有数据集:来自莱顿大学医学中心,包含 288 名不同患者的 792 个 CT 扫描	CT 图像(3D)	食管肿瘤分割(GTV)	扩张卷积增大感受野但降低特征分辨率;密集连接增加内存消耗;注意力机制在低信噪比区域易受干扰
SOARS(Stratified Organ at Risk Segmentation)新型分层学习框架,使用神经架构搜索(NAS)为每类器官自动选择最佳网络结构,使用 P-HNN(Progressive and Multi-path Holistically Nested Neural Networks)作为主干网络 <sup>[36]</sup>	私有数据集:H&N 42 OARs 数据集,包含 142 个匿名非对比增强放射治疗 CT 图像,每个扫描包含 42 个器官风险(OARs)的 3D 掩码标注公开数据集: MICCAI 2015 头颈部自动分割挑战赛数据集,包含 33 个训练病例,15 个测试病例,9 个 OARs 标注	CT 图像(3D)	头颈部癌症放射治疗规划中的器官风险(OARs)自动分割	分层学习框架需手动定义器官风险等级,主观性强;NAS(神经架构搜索)计算成本高;模型对罕见器官分割性能不稳定
“You-Only-Have-One”(YOHO) 框架(CNN+Single-shot) <sup>[37]</sup>	私有数据集:EEC-2022 数据集,作者自行创建的早期食管癌数据集,包含 1 236 张图像公开数据集: CVC-ClinicDB 数据集;Kvasir-SEG 数据集	内窥镜图像包括食管早期癌症和结肠息肉图像(2D)	早期食管癌(EEC)病变和结肠息肉的分割	单张图像输入缺乏时序或空间上下文信息;模型对光照变化和黏膜反光敏感;结肠息肉与食管癌数据混合训练导致特征混淆
基于病变解耦的分割(Lesion-Decoupling-based Segmentation, LDS)网络(自采样相似特征解耦模块 FDM+特征分离损失函数 FSL+CNN) <sup>[38]</sup>	私有数据集:大规模结肠镜数据集: 28 485 张图像,食管镜多模态配对数据集,454 对 WLI 和 NBI 图像公开数据集: ETIS Larib Polyp DB, CVC-ClinicDB, Kvasir-SEG	结肠镜白光图像;食管镜白光图(WLI);食管镜窄带成像(NBI)(2D)	结肠癌和食管癌早期诊断;结肠息肉和食管病变区域检测和分割	病变解耦模块依赖人工定义特征分离准则;多模态数据(WLI/NBI)配准误差影响模型性能;特征分离损失函数难以平衡分割精度与泛化性
形状感知对比深度监督网络 SCDS-Net(Shape-aware Contrastive Deep Supervision Network) <sup>[39]</sup>	私有数据集:收集自山东省肿瘤医院,558 例食道癌患者的胸部 CT 扫描	CT 扫描(3D)	食道肿瘤分割	形状感知正则化对肿瘤形态先验知识依赖性强;对比学习需大量负样本,实际标注成本高;3D 监督信号稀疏,训练收敛速度慢
结合多种骨干网络的 U-Net 神经网络,包括 ResNet, SE-ResNet, ResNeXt, InceptionNet 等 <sup>[40]</sup>	私有数据集:来自山东省肿瘤医院放射肿瘤科,共 17 次扫描,总计 1 397 张切片	CT 图像(2D)	计算机断层扫描(CT)图像中食管分割	多骨干网络参数量较大,训练时间倍增;特征融合策略简单,未充分利用多尺度信息;水平集算法生成边缘依赖人工初始化,自动化受限
多任务卷积神经网络,包括下采样路径:预训练的 ResNet;上采样路径:通过进化神经架构搜索优化 <sup>[41]</sup>	私有数据集:来自四川大学华西医院 549 名患者,包含 6 473 张 NBI 食管图像	窄带成像(NBI)内窥镜图像(2D)	自动食管病变分类和分割	进化算法搜索空间较大,超网络策略仍需多次迭代;多任务联合训练中分类与分割目标冲突;模型对窄带成像(NBI)的血管模式变化敏感,泛化性不足

CNN 技术在 EC 早期筛查和诊断中展现了重要应用价值。Liu 等<sup>[22]</sup>提出了一种基于深度学习的 DD-BA 方法,用于自动标注胃镜图像中的早期 EC 病灶,该方法包含 4 个步骤:去除背景,预处理生成深度图,结合 RGB 图像预测病灶区域,进行后处理。该方法结合深度与 RGB 信息显著提高了标注精度,其误报率低于其他方法。鉴于 CNN 网络可解释性较低,Shaw 等<sup>[5]</sup>通过 LIME 解释性机器学习技术建立信任模型,利用简单 CNN 和迁移学习模型检测早期 EC,并在 Kvasir-v2 数据集上进行实验,DenseNet-201 模型达到了 88.75% 的准确率,LIME 解释了模型的关键区域,增强了模型可信度。Fu 等<sup>[23]</sup>为降低 YOLO-V5 在食管病变检测中的复杂性,提出了基于极化正则化和网络剪枝的 EPS YOLO-V5。该模型结合 GhostNet 模块构建特征提取网络,通过稀疏性和正则化优化 BN 层,进行自动剪枝并微调,最终实现参数减少 5 倍、模型大小减少 8 倍、推理速度提高 4.5 倍,同时保持与原 YOLO-V5 相当的检测性能。针对医疗场景下数据有限性的问题,Yan 等<sup>[24]</sup>提出了自监督特征表示蒸馏(Self-supervised Feature Representation Distillation, SFRD)方法。该方法通过教师模型与学生模型相结合,利用自监督学习与知识蒸馏增强特征提取能力,实验结果表明,SFRD 在临床数据集上优于其他方法。综上所述,CNN 高效的图像特征提取能力使其在 EC 诊疗中展现出发展潜力<sup>[25]</sup>。

CNN 技术在 EC 疗效预测和生存分析中也展现出重要的价值。EC 生存预测对于个性化治疗至关重要,传统预测方法依赖人工特征提取,需依靠先验医学知识,限制了预测的效率和精度。针对此类问题,Lin 等<sup>[30]</sup>提出了一种基于深度学习的生存预测框架,用于同步放化疗前的临床结果评估。该框架结合 3D 坐标注意力卷积自编码器(Coordinate Attention Convolutional Autoencoder, CACA)和不确定性联合优化损失函数,用于提高捕获空间特征的精确度。该方法在 285 名 EC 患者的 CT 数据集上取得了 0.72 的 C-index 指数。Guan 等<sup>[21]</sup>提出将 CT 图像和临床数据相结合的方法,利用自注意力模块 VisText 揭示不同特征间的复杂关系,通过聚合多模态特征提升了预测准确性。然而,现有研究大多仅关注原发肿瘤 CT 图像,忽略了淋巴结在预后中的作用。针对此问题,Wu 等<sup>[31]</sup>提出多模态人群图框架,整合原发肿瘤与淋巴结 CT 图像以及临床、血液学、放射组学数据,以挖掘特征同质性和异质性,用于 EC 预后风险预测。实验结果证明该方法具有较强的优越性。

CNN 技术在 EC 分割中展现了重要应用价值。由于食管区域比例小且形状不规则,准确进行 EC 的分割具有较大挑战。Huang 等<sup>[32]</sup>提出通道注意力 U-Net,结合通道注意力模块(Channel Attention Module, CAM)和跨层特征融合模块(Cross-layer Feature Fusion Module, CFFM),强调有效特征的提取并增强泛化能力,取得了 0.732 的 Dice 系数和 3.193 的 Hausdorff 距离。Zhou 等<sup>[33]</sup>提出 Eso-Net,结合多结构响应滤波器(Multi-structure Response Filter, MSRF)、空洞卷积、残差连接和先验注意力机制模块(Prior-attention Module, PAM),在 30 例 EC 患者数据集上进行测试,获得了 84.8% 的 Dice 系数,优于现有方法。Sun 等<sup>[34]</sup>设计了基于 3D U-

Net 的区域保留注意力模块(Region-preserving Attention Module, RAM),有效保留肿瘤特征,提高了二次放疗中 GTV 分割的准确性。Yousefi 等<sup>[35]</sup>提出扩张密集注意力 U-Net(DDAUnet),利用空间和通道注意力聚焦关键特征,同时通过空洞卷积扩大感受野,取得了良好的临床效果。Guo 等<sup>[36]</sup>提出 SOARS 分割方法,用于放疗中的危险器官(Organ at Risk, OAR)分割,在有 142 名患者的数据集上进行测试,Dice 系数从 70.44% 提升至 75.14%。

对于早期 EC 的诊断与治疗,精准的病灶分割至关重要。Li 等<sup>[37]</sup>提出“You-Only-Have-One”(YOHO)框架,该方法仅依赖单张图像,避免泛化问题,平均 Dice 系数达到 0.888,显著高于现有方法。Lin 等<sup>[38]</sup>提出病变解耦分割网络(Lesion Decoupled Segmentation, LDS),通过自采样相似特征解耦模块(Feature Decoupling Module, FDM)和特征分离损失(Feature Separation Loss, FSL)来提高病变分割的准确率,在食管镜检查中,WLI 和 NBI 数据集上的 mIoU 分别为 64.32 和 66.31。Jin 等<sup>[39]</sup>提出了形状感知对比深度监督网络(SCDS-Net),结合形状感知正则化和对比监督,在大规模 EC 数据集上展现出优异效果。这些研究表明,CNN 技术在提高食管癌分割的准确性和治疗效果方面具有较大潜力。

CNN 技术在 EC 多任务诊疗模型中也展现了重要价值。多任务机器学习模型通常集成分割、分期分类和预后预测于一体,旨在提高诊疗效率和准确性。Lou 等<sup>[40]</sup>提出了一种基于多骨干网络的 U-Net 模型,用于食管半自动标注和分割。该方法通过特征袋、轮廓矩、凸包和水平集算法生成病灶边缘并进行训练,其中以 ResNeXt50 和 InceptionV3 作为骨干网络时效果最佳。Zhou 等<sup>[41]</sup>提出一种基于进化算法(EA)的优化方法,用于搜索 ESEI 的最优多任务网络架构,该方法通过一次性超网络策略提高效率,避免了原 EA 中的重新训练,实验结果显示,其在各项指标上均优于现有方法。

CNN 在单模态影像分析中表现卓越,但全局上下文建模能力有限。在生存预测中,Lin 等<sup>[30]</sup>仅依赖 CT 影像,其 C-index 显著低于多模态方法。多模态 CNN 通过跨模态交互进一步提升了性能。Wang 等<sup>[42]</sup>提出多模态 Transformer,通过跨注意力机制整合 CT 与病理报告,将分割 Dice 提升至 81.2%(单模态 75.1%)。Wu 等<sup>[31]</sup>利用图神经网络(GNN)融合 CT 影像与血液指标,C-index 达 0.83(单模态 0.72);Chen 等<sup>[43]</sup>利用对比学习对齐影像与基因组数据,使分类准确率提升 12%。尽管多模态 CNN 显著优化了模型性能,但计算复杂度高且依赖人工定义模态关联。

CNN 尽管在 EC 的临床诊疗中展现了显著的应用价值,但其在长程依赖建模方面的局限性,导致其难以有效捕捉全局信息和长程依赖关系。尤其在处理具有复杂形态学特征的医学图像时,仅仅关注局部特征难以满足模型对特征表征的全面需求,模型需要对特征的全局变化进行更精确的建模。此外,CNN 作为一种“黑箱”模型,其决策过程缺乏透明性和可解释性<sup>[44]</sup>,而这点在医学领域尤为关键,直接影响着临床决策的可靠性和依据性。这些局限性显著限制了 CNN 在 EC 诊疗中的进一步推广和应用。

### 3.3 CNN 和 Transformer 结合技术在 EC 诊疗中的应用

CNN 擅长捕获局部特征,适合处理空间信息,有效降低计算成本,而 Transformer<sup>[9]</sup> 在捕捉数据中的长距离依赖性方面表现出色。因此,将二者结合有助于全面理解图

像内容,尤其是细节和全局上下文之间的关联。这种融合网络在 EC 的早期筛查和诊断、疗效预测和生存分析、EC 分割、EC 多任务诊疗中表现出了较大的诊断价值,如表 3 所列。

表 3 CNN 和 Transformer 结合技术在 EC 中的应用

Table 3 Application of CNN and Transformer combination technology in EC

网络类型	数据来源	数据类型	临床任务	缺陷和不足
EsceNet, 混合 CNN 和 Transformer 模型 <sup>[47]</sup>	私有数据集:来自福建省肿瘤医院 50 个病例,包含 73 张食管鳞状细胞癌全切片图像(WSI)	食管鳞状细胞癌全切片图像(WSI)(2D)	食管鳞状细胞癌全切片图像的四分类	Transformer 模块计算复杂度高,处理大尺寸 WSI 时内存消耗较大;CNN 与 Transformer 特征融合策略简单,未充分挖掘跨尺度信息;模型对染色差异敏感
CNN-Transformer 结合的混合网络架构 <sup>[48]</sup>	私有数据集:来自西京医院,共 759 名接受 dCRT 治疗的食管癌患者	CT 图像(3D)	食管癌患者生存预测;自动分割感兴趣区域(ROI)预测患者生存风险	多任务训练中生存预测与分割目标冲突;ROI 分割依赖人工标注,自动化程度低;Transformer 模块在 3D 数据中远程依赖建模效率低
基于自编码器的深度学习模型,多模态中间特征交互自编码器(MIFI-AE)(包含编码器和解码器部分,使用残差块和 Swin Transformer) <sup>[6]</sup>	私有数据集:来自四川省肿瘤医院,共 1354 名食管鳞状细胞癌(ESCC)患者	术前 CT 图像和肿瘤靶区体积(GTV-T)(3D) 临床血液学参数和随访信息(2D)	食管鳞状细胞癌(ESCC)患者总生存期预测	模态间交互机制参数量较大,训练数据不足时易过拟合;未考虑时序临床数据的动态影响;跨模态对齐依赖先验知识,泛化性差
基于 TransUNet 改进的 CCA-TransUNet <sup>[8]</sup>	私有数据集:来自兰州大学第一医院放疗科,共 17 例食管癌 CT 图像,超过 1400 个切片	CT 图像(2D)	食管癌肿瘤分割	双通道注意力模块增加计算冗余;3D 重建依赖切片拼接,空间连续性建模不足;对低对比度肿瘤边界分割效果不稳定
基于 nnUNet 的改进网络架构,在编码器的卷积层之后添加了位置敏感的自注意力层(Position-sensitive Self-attention) <sup>[49]</sup>	私有数据集:来自 741 名患者的 CT 扫描,包含 481 例有食管肿瘤(癌症或良性),其中 260 例为正常病例	CT 图像(3D)	食管肿瘤检测,肿瘤分类和早期筛查	自注意力模块引入额外计算开销,实时性差;位置编码对非刚性解剖结构适应不足;多中心数据分布差异导致模型性能波动

在 EC 早期筛查和诊断中,Kang 等<sup>[47]</sup> 提出了一种创新方法,结合 CNN 和 Transformer 处理全病理切片图像(Whole Slide Image, WSI),用于食管鳞状细胞癌(Esophageal Squamous Cell Cancer, ESCC)分类。该方法利用 CNN 提取局部特征,通过 Transformer 处理全局背景,最终实现 94.71% 的准确率,显著优于其他模型,为病理分析和癌症诊断提供了重要参考。在疗效预测和生存分析中,CNN 与 Transformer 的结合同样展现出应用前景。Yue 等<sup>[48]</sup> 提出了 CNN-Transformer 结合的大小感知 U-Net 与排序生存网络,用于 CT 引导的 EC 生存预测。该方法通过移动窗口多头自注意力捕捉 CT 图像的长程依赖性,结合大小感知系数解决 ROI 与背景不平衡问题,提升了生存预测的准确性。此外,Wu 等<sup>[6]</sup> 提出了一种基于自编码器的多模态深度学习模型,旨在更好地捕捉跨模态预后特征和语义差异,实验结果表明,该模型在区分能力和风险分层方面表现优越。在分割任务中,CNN 与 Transformer 的结合也表现出了潜力。Yu 等<sup>[8]</sup> 提出 CCA-TransUNet,将传统 U-Net 与 Transformer 相结合,引入双通道高效注意力(Dual-channel Efficient Attention, DEA)模块,以整合坐标通道注意力(Coordinate Channel Attention, CCA)和高效通道注意力(Efficient Channel Attention, ECA)。该方法在食管 CT 分割中实现了 76.44% 的平均 Dice 相似系数和 3.45mm 的 Hausdorff 距离,显著提高了分割性能。

在多任务诊疗中,CNN 与 Transformer 的结合同样表现优异。Yao 等<sup>[49]</sup> 提出了基于 Transformer 的位置敏感自注意力(Position Sensitive Self Attention, PSSA)模块,以增强 nnUNet 的非局部交互能力。在 180 名患者的测试中,模型在肿瘤检测中的敏感性为 93.0%,特异性为 97.5%,均显著高

于医生的平均水平。此外,该模型在癌症、良性肿瘤和正常组织分类中的表现也明显优于传统手段,且该方法非侵入、低成本,更适用于早期癌症和良性肿瘤的筛查。总的来说,CNN 与 Transformer 结合在食管癌的筛查、诊断、分割及多任务诊疗中表现出广泛的应用价值,为未来个性化诊疗提供了新的方向和可能性。

与单模态全局建模相比,多模态交互成为混合模型的核心优势。Wu 等<sup>[6]</sup> 提出了 MIFI-AE 模型,利用 Swin Transformer 融合 CT 影像与基因组数据,生存预测的 C-index 达 0.81(单模态 0.72)。Zhang 等<sup>[50]</sup> 通过 GNN 关联 MRI 与临床数据,分类 AUC 提升至 0.91(单模态 0.82)。Li 等<sup>[51]</sup> 采用联邦学习实现跨中心多模态模型训练,分割 Dice 系数提升 8% 且满足隐私合规要求,但模态间参数冗余与实时性不足仍是当前的瓶颈。当前多模态融合在 EC 诊疗中面临模态异构性、计算效率及可解释性 3 大挑战。影像与文本的语义差异需利用更精细的对齐策略,如对比学习。而 Transformer 与 GNN 的高复杂度限制了实时应用,需结合知识蒸馏<sup>[24]</sup> 等压缩技术。未来方向包括设计轻量级架构,如线性注意力、联邦学习支持下的跨中心协作<sup>[51]</sup>,以及结合因果推理探索模态间关联<sup>[52]</sup>。

尽管结合 CNN 和 Transformer 的混合模型在 EC 诊疗中展现出较大的应用潜力,但相关研究尚处于起步阶段,该类模型仍存在以下不足。1) 特征提取能力不足:现有模型在有效提取关键特征方面仍存在局限性,多模态数据的引入未能显著提升模型性能,甚至部分情况下未显示出提升效果。这主要归因于模型缺乏针对冗余信息的有效过滤机制,导致在生存分析任务中,C-index 指标普遍偏低,无法满足临床的需

求<sup>[53]</sup>。2)多模态数据交互融合不足:当前模型在CT图像与临床文本数据的融合方面缺乏统一且高效的交互机制,未能实现多模态信息的深度协同与互补,这种局限性抑制了不同模态数据的有效整合,进而对模型的整体性能产生负面影响<sup>[6]</sup>。3)多任务协同不足:现有模型大多采用独立处理任务的方式,未能充分挖掘任务间的互补性与共享信息。这种单一任务处理模式限制了模型在生存分析与病灶分割等任务中的协同效应,无法达到系统级诊断性能的最优表现<sup>[54]</sup>。因此,为进一步提升智能诊断模型的效果,针对特征提取、多模态数据交互融合以及多任务协同机制的改进与优化显得尤为关键。这将有助于突破当前模型的性能瓶颈,为EC的精准诊疗提供更加可靠的技术支持。

### 3.4 机器学习在食管癌诊疗中的可解释性研究

机器学习模型的可解释性在食管癌的诊疗中具有重要的作用。医学决策通常关系到患者的健康和生命安全,因此,医生和患者需要了解模型做出预测或诊断的具体原因。可解释性有助于增强模型的可信度,尤其在食管癌的自动诊断、治疗方案推荐和疾病风险预测中,当模型输出与临床经验不一致时,具备可解释性的模型能够提供透明的依据,帮助医疗人员判断结果的可靠性。在食管癌的诊疗中,通常通过下面3种策略来增加模型的可解释性。

第一种策略是借助于另一种具有可解释性的辅助模型,增加预测模型的关键环节的可解释性。例如LIME(Local Interpretable Model-agnostic Explanations)模型,通过扰动输入数据生成局部线性模型,可视化CNN对食管癌病灶区域(如边缘不规则性、血管异常)的关注权重。在文献<sup>[5]</sup>中,LIME的解释结果与病理学家的标注一致性达85%,显著提升了医生对AI的信任度。

第二种策略是借助于自注意力机制等模型较强的可解释性,增加模型逻辑推理的可视化程度。如利用Transformer的自注意力得分图,揭示模型对图像等重点关注区域。例如,Wu等<sup>[6]</sup>通过跨模态注意力热图得出,生存预测模型主要依赖肿瘤体积(GTV-T)和淋巴细胞计数,具有较高的可解释性,增加了模型的可信度。

第三种策略是对特征的重要性进行可视化,增加模型的可解释性。这种方法在传统机器学习中具有较多的应用。例如可通过随机森林和SHAP(SHapley Additive exPlanations)来量化临床参数(如TNM分期、CEA水平)对预后的贡献度,以增加模型的可解释性。在文献<sup>[16]</sup>中,SHAP分析显示,术前化疗周期数是复发预测中的最关键特征(贡献度32%),可以帮助医务人员确定复发的危险因素,同时可以和医疗经验进行更好的互相验证,提高了模型的可信度和可推广性。

随着机器学习研究的发展,深度学习模型虽然在医学中的可解释性得到了显著提高,但是仍然面临着诸多挑战。其中,模型结构复杂度的提高使得诊断预测等任务的可解释性变得更加困难,这导致医生难以信任模型输出,模型的“黑箱”性质可能影响医疗决策的安全性和可靠性。因此,如何进一步提高模型的可解释性仍然是食管癌智能诊疗能否真正在临床普及落地的一大挑战。

## 4 机器学习方法对比与临床适用性分析

如何提高食管癌的生存分析以及食管癌CT图像的分割效果,结合前期学者的研究工作和笔者自身的经验,以下几个方面的经验可以供读者参考。1)数据模态的重要性。CT图像在食管癌生存分析预测中具有重要的作用,相比于X射线图像,CT包含更加丰富的生存信息。未来MR,PET等反映组织代谢的数据,以及更有价值的分子标志物的加入必将使得食管癌的生存分析预测水平更上一个台阶。2)多模态信息的应用相比于单模态,能进一步提升模型的预测效果。在EC生存分析中,临床文本数据记录患者病历、检验结果及治疗信息,反映病史及风险;CT图像数据则提供肿瘤形态、大小及边缘等特征,揭示肿瘤进展情况。两者融合后,优势互补,提高了生存预测模型的准确性,为精准医疗提供全面依据,具体可参考文献<sup>[6]</sup>。多模态融合需解决模态对齐问题,如文献<sup>[28]</sup>通过自注意力模块关联CT与临床文本,但计算成本较高。3)低计算复杂度模型在食管癌的临床诊疗中具有重要的价值。低计算复杂度模型运算速度更快、数据需求量更小,模型结构简单、易于解释。例如,基于线性注意力机制的模型在食管癌的诊断中具有更大的应用潜力<sup>[55]</sup>。在计算效率与适用场景上,传统ML(如SVM和随机森林)参数少、训练快,适用于小规模结构化数据(如血液指标),但其特征工程依赖领域知识;CNN通过局部卷积核提取图像特征,使得图像分割中实时推理速度提高了4.5倍<sup>[18]</sup>,但需大规模标注数据;Transformer因自注意力机制的计算复杂度( $O(n^2)$ )受限,Linformer<sup>[42]</sup>通过低秩近似将复杂度降至 $O(n)$ ,更适合长序列或多模态数据。4)矫正食管癌CT图像中的类别不平衡和长尾问题,对于提高食管癌生存分析的预测效果和食管癌CT图像的分割效果具有重要的作用。在关键参数调试与训练中,CNN需优化卷积核尺寸、池化策略及正则化强度,以平衡过拟合与泛化性;Transformer需调整注意力头数、位置编码方式及投影维度,临床应用中常因数据稀缺而需结合迁移学习<sup>[19]</sup>。不同的模型可以采用不同的策略进行矫正,且都取得了较好的预测结果,具体可参考文献<sup>[56]</sup>。

## 5 总结

### 5.1 机器学习在EC诊疗中存在的问题

尽管机器学习技术在EC的诊疗中展现了显著的应用价值,但其仍然有很多问题需要解决。目前,机器学习模型在EC的二分类鉴别诊断中表现出较高的准确性( $AUC > 0.85$ ),但在时间相关事件的预测以及EC分割中的表现仍然不佳。例如,在生存分析中,C-index通常低于0.72,在EC的分割中,Dice系数通常在0.8以下,均未达到临床要求。其原因主要有3点。1)数据模态不匹配:对近5年EC研究的文献分析显示,约70%的研究基于病理、内镜和钡餐造影图像,这些图像虽然对诊断有较大价值,但在疗效预测中却无法提供足够的信息量,无法为生存分析提供足够信息支持。2)EC图像的类别不平衡现象严重:与体部其他肿瘤相比,EC在整幅图像中所占的比例通常较小,且其形态不规则,这给EC的分割和信息提取带来了较大挑战。3)多模态数据融合效果

差:生存分析和分割的准确性依赖于多模态数据的深度融合,但目前大多数的研究通常仅将不同模态数据简单组合,难以有效提取跨模态的关键信息来促进不同任务之间的协同作用,限制了模型的性能。

## 5.2 EC 机器学习模型在临床转化中存在的问题

首先,通过对近 5 年的研究结果进行分析发现,超过半数的模型仅在构建的私有数据集上完成了训练和测试。由于不同医疗中心的数据存在显著差异,模型能否在其他数据集上保持较高的预测和分割准确性,仍需进一步研究和验证。解决这一问题不仅需要扩展数据集的多样性,还需通过跨中心训练和测试来提升模型的泛化能力。其次,技术迭代和维护成本也是模型走向临床的一大挑战。单中心模型的性能可能随着时间推移而下降(数据漂移),在多中心环境中进行持续训练和更新是确保模型长期有效的关键。然而,这一过程往往需要耗费大量资源,如何在成本可控的情况下实现模型的技术迭代和维护,是临床应用中需要重点关注的问题。最后,法律与伦理合规性问题也是制约模型临床转化的重要因素。单中心数据的隐私保护和伦理审批通常仅限于数据采集中心,然而模型的临床应用需要跨越多中心进行数据共享和验证。这要求在充分保护患者隐私的前提下,严格遵守各地的法律法规。如何收集和利用多中心数据对模型进行训练和测试,同时兼顾成本、隐私保护以及合规性,是影响模型临床化的重要瓶颈之一。综上所述,机器学习模型在目前的数据集上虽已表现出了较高的准确性,但要真正实现临床落地,还需解决数据多样性、技术迭代、隐私保护与法规遵从等多重难题,这也是未来研究和开发工作的重点方向。

## 5.3 EC 机器学习在临床转化中的未来趋势和热点

尽管机器学习技术在 EC 的诊疗中展现了显著的应用潜力,但要真正实现其在临床中的转化,仍需克服 3 个方面的挑战。首先,优化算法和模型性能是关键。通过持续改进模型算法、探索无监督或弱监督建模技术<sup>[11]</sup>,并结合图神经网络<sup>[57]</sup>和经典图像处理方法,能够增强模型对多模态数据的处理能力,提高其临床应用价值。其次,扩大数据规模和模态种类至关重要。多中心、多模态的数据来源有助于提升模型的泛化能力和精确度,尤其是在引入 MRI、内镜图像等新型数据的基础上,推动跨模态特征的提取与关联机制的发展。此外,聚焦临床应用场景,针对不同复杂度的诊疗需求,设计灵活高效的智能诊断工具,以提升模型的实用性和医生的接受度。

根据目前领域的研究热点,未来的研究可进一步聚焦于以下 3 方面。1)临床转化瓶颈:模型泛化性受限于单中心数据,未来需通过联邦学习实现跨中心协作。2)实时性与可解释性:开发轻量级模型(如 EPS YOLO-V5<sup>[23]</sup>)并结合 LIME<sup>[5]</sup>等可解释技术,提升医生信任度。3)多模态深度融合:探索图神经网络<sup>[57]</sup>与知识图谱,关联影像、基因组与电子病历数据。未来研究应进一步聚焦个性化诊疗系统,推动 EC 诊疗从“经验驱动”向“数据驱动”的范式转变。

总之,通过在算法优化、数据扩展和临床应用场景等方面的持续改进,使机器学习模型最终转化为可依赖的智能诊断工具,为 EC 患者的精准治疗提供支持。

## 参 考 文 献

- [1] SUNG H, FERLAY J, SIEGEL R L, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries [J]. *CA Cancer J Clin*, 2021, 71(3): 209-249.
- [2] QIU H, CAO S, XU R. Analysis of the time trend of China's cancer incidence rate, mortality and burden based on the global epidemiological data in 2020 and comparison with data from the United States and the United Kingdom [J]. *Cancer*, 2022, 41(4): 165-177.
- [3] DENG W, LIN S H. Advances in radiotherapy for esophageal cancer [J]. *Annals of Translational Medicine*, 2018, 6(4): 79.
- [4] KATO H, NAKAJIMA M. Treatments for esophageal cancer: a review [J]. *General Thoracic and Cardiovascular Surgery*, 2013, 61: 330-335.
- [5] SHAW P, SANKARANARAYANAN S, LORENZ P. Early Esophageal Malignancy Detection Using Deep Transfer Learning and Explainable AI [C] // Proceedings of the 6th International Conference on Communication and Information Systems. IEEE, 2022.
- [6] WU C, ZHANG Y, WANG Y, et al. Multi-Modal Intermediate Feature Interaction Autoencoder for Overall Survival Prediction of Esophageal Squamous Cell Cancer [C] // Proceedings of the 2024 IEEE International Symposium on Biomedical Imaging. IEEE, 2024.
- [7] LI Z, LIU F, YANG W, et al. A survey of convolutional neural networks: analysis, applications, and prospects [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 33(12): 6999-7019.
- [8] YU Z, JIANG S, WANG L. Esophageal Image Segmentation with Dual Attention Based on TransUNet [C] // Proceedings of the 2024 4th International Conference on Consumer Electronics and Computer Engineering. IEEE, 2024.
- [9] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 6000-6010.
- [10] WANG S, LI B Z, KHABSA M, et al. Linformer: Self-Attention with Linear Complexity [J]. *arXiv*: 2006. 04768, 2020.
- [11] AZAD R, KAZEROUNI A, HEIDARI M, et al. Advances in medical image analysis with vision transformers: a comprehensive review [J]. *Medical Image Analysis*, 2024, 91: 103000.
- [12] ERICKSON B J, KORFIATIS P, AKKUS Z, et al. Machine learning for medical imaging [J]. *Radiographics: A Review Publication of the Radiological Society of North America*, 2017, 37(2): 505-515.
- [13] YAMAGUCHI J, YONEYAMA A, MINAMOTO T. Automatic detection of early esophageal cancer from endoscope image using fractal dimension and discrete wavelet transform [C] // Proceedings of the 12th International Conference on Information Technology-New Generations. IEEE, 2015.
- [14] XUE Y, LI N, WEI X, et al. Deep learning-based earlier detection of esophageal cancer using improved empirical wavelet

- transform from endoscopic image [J]. *IEEE Access*, 2020, 8: 123765-123772.
- [15] PADHA A, SAHOO A. Quantum deep neural networks for time series analysis [J]. *Quantum Information Processing*, 2024, 23(6):205.
- [16] LYU H, KAPCIO K, PURRMAN K, et al. Applying Machine Learning to Predict Esophageal Cancer Recurrence after Esophagectomy[C] // *Proceedings of the 2023 IEEE International Conference on Digital Health*. IEEE, 2023.
- [17] SU Y, HUANG C, YANG C, et al. Prediction of Survival in Patients With Esophageal Cancer After Immunotherapy Based on Small-Size Follow-Up Data [J]. *IEEE Open Journal of Engineering in Medicine and Biology*, 2024(5):769-782.
- [18] SUN J, YANG Y, WANG Y, et al. Survival risk prediction of esophageal cancer based on self-organizing maps clustering and support vector machine ensembles [J]. *IEEE Access*, 2020, 8: 131449-131460.
- [19] PRINCE T, WONDMANEHGETAHUN B, AMBACHEW-GOSHU K, et al. Multi-Classification and Segmentation of Esophageal Lesions Using an Improved Deep Learning Model from Endoscopic Images[C] // *Proceedings of the 2023 Eighth International Conference on Science Technology Engineering and Mathematics*. IEEE, 2023.
- [20] NGIAM J, KHOSLA A, KIM M, et al. Multimodal Deep Learning[C] // *Proceedings of the International Conference on Machine Learning*. IEEE, 2011.
- [21] GUAN Y, CUI H, XU Y, et al. Predicting esophageal fistula risks using a multimodal self-attention network[C] // *Proceedings of the Medical Image Computing and Computer Assisted Intervention*, 2021.
- [22] LIU D, JIANG H, RAO N, et al. Depth information-based automatic annotation of early esophageal cancers in gastroscopic images using deep learning techniques [J]. *IEEE Access*, 2020, 8:97907-90919.
- [23] FU Y, ZHOU Y, YUAN X, et al. Efficient Esophageal Lesion Detection using Polarization Regularized Network Slimming [C] // *Proceedings of the 2022 IEEE 8th International Conference on Cloud Computing and Intelligent Systems*. IEEE, 2022.
- [24] YAN X, ZHOU Y, YI Z. Self-supervised Feature Representation Distillation for Esophageal Cancer Screening[C] // *Proceedings of the 2024 International Joint Conference on Neural Networks*. IEEE, 2024.
- [25] VAN RIEL S, VAN DER SOMMEN F, ZINGER S, et al. Automatic detection of early esophageal cancer with CNNs using transfer learning[C] // *Proceedings of the 2018 25th IEEE International Conference on Image Processing*. IEEE, 2018.
- [26] ABURASAIN R Y. Esophageal Cancer Classification in Initial Stages Using Deep and Transfer Learning[C] // *Proceedings of the 2024 IEEE International Conference on Advanced Systems and Emergent Technologies*. IEEE, 2024.
- [27] TSAO Y M, MUKUNDAN A, KARMAKAR R, et al. Hyperspectral Imaging Applied to Identify Early Esophageal Cancer [C] // *Proceedings of the 2024 Conference on Lasers and Electro-Optics Pacific Rim*. IEEE, 2024.
- [28] DING S, HUANG H, LI Z, et al. SCNET: A novel UGI cancer screening framework based on semantic-level multimodal data fusion [J]. *IEEE Journal of Biomedical and Health Informatics*, 2020, 25(1):143-151.
- [29] GHATWARY N, ZOLGHARNI M, JANAN F, et al. Learning spatiotemporal features for esophageal abnormality detection from endoscopic videos [J]. *IEEE Journal of Biomedical and Health Informatics*, 2020, 25(1):131-142.
- [30] LIN Z, CAI W, HOU W, et al. CT-guided survival prediction of esophageal cancer [J]. *IEEE Journal of Biomedical and Health Informatics*, 2021, 26(6):2660-2669.
- [31] WU C, WANG S, WANG Y, et al. A Novel Multi-modal Population-graph based Framework for Patients of Esophageal Squamous Cell Cancer Prognostic Risk Prediction [J]. *IEEE Journal of Biomedical and Health Informatics*, 2024, 29(5):3206-3219.
- [32] HUANG G, ZHU J, LI J, et al. Channel-attention U-Net: Channel attention mechanism for semantic segmentation of esophagus and esophageal cancer [J]. *IEEE Access*, 2020, 8: 122798-122810.
- [33] ZHOU D, HUANG G, LI J, et al. Eso-net: A novel 2.5D segmentation network with the multi-structure response filter for the cancerous esophagus [J]. *IEEE Access*, 2020, 8: 155548-155562.
- [34] SUN Y, KHOR H G, HUANG S, et al. Second-Course Esophageal Gross Tumor Volume Segmentation in CT with Prior Anatomical and Radiotherapy Information[C] // *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. IEEE, 2023.
- [35] YOUSEFI S, SOKOOTI H, ELMAHDY M S, et al. Esophageal tumor segmentation in CT images using a dilated dense attention Unet(DDAUnet) [J]. *IEEE Access*, 2021, 9:99235-99248.
- [36] GUO D, JIN D, ZHU Z, et al. Organ at risk segmentation for head and neck cancer using stratified learning and neural architecture search[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
- [37] LI H, LIU D, ZENG Y, et al. Single-Image-Based Deep Learning for Segmentation of Early Esophageal Cancer Lesions [J]. *arXiv*:2306.05912, 2023.
- [38] LIN Q, TAN W, CAI S, et al. Lesion-decoupling-based segmentation with large-scale colon and esophageal datasets for early cancer diagnosis [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 35(8):11142-11156.
- [39] JIN Q, CUI H, SUN C, et al. Shape-aware contrastive deep supervision for esophageal tumor segmentation from CT scans [C] // *Proceedings of the 2023 IEEE International Conference on Bioinformatics and Biomedicine*. IEEE, 2023.
- [40] LOU X, ZHU Y, PUNITHAKUMAR K, et al. Esophagus segmentation in computed tomography images using a U-Net neural network with a semiautomatic labeling method [J]. *IEEE Access*, 2020, 8:202459-202468.
- [41] ZHOU Y, YUAN X, ZHANG X, et al. Evolutionary neural architecture search for automatic esophageal lesion identification and segmentation [J]. *IEEE Transactions on Artificial Intelligence*, 2021, 3(3):436-450.

- [42] WANG S, LI B Z, KHABSA M, et al. Linformer: Self-attention with linear complexity [J]. arXiv:2006.04768, 2020.
- [43] CHEN T, KORNBLITH S, NOROUZI M, et al. A Simple Framework for Contrastive Learning of Visual Representations [J]. arXiv:2002.05709, 2020.
- [44] RAJKOMAR A, DEAN J, KOHANE I S. Machine Learning in Medicine [J]. The New England Journal of Medicine, 2019, 380: 1347-1358.
- [45] TJOA E, GUAN C. A Survey on Explainable Artificial Intelligence(XAI): Toward Medical XAI [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 32: 4793-4813.
- [46] SAMEK W, MONTAVON G, VEDALDI A, et al. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning [M]. Cham: Springer, 2019.
- [47] KANG Z, ZHANG H, CHEN M, et al. EsceNet: A Hybrid CNN and Transformers Model for the Classification of Whole Slide Images of Esophageal Squamous Cell Carcinoma [C] // Proceedings of the 2024 5th International Conference on Computer Engineering and Application. IEEE, 2024.
- [48] YUE H, LIU J, KUANG H, et al. A Fully Automated CT-Guided Learning for Survival Prediction of Esophageal Cancer [C] // 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2023: 1670-1675.
- [49] YAO J, YE X, XIA Y, et al. Effective opportunistic esophageal cancer screening using noncontrast CT imaging [C] // Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2022.
- [50] ZHANG Y, HE N, YANG J, et al. mmFormer: Multimodal Medical Transformer for Incomplete Multimodal Learning of Brain Tumor Segmentation [C] // Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. 2022.
- [51] LI T, SAHU A K, TALWALKAR A, et al. Federated Learning: Challenges, Methods, and Future Directions [J]. IEEE Signal Processing Magazine, 2019, 37: 50-60.
- [52] FRAPPIER M. The Book of Why: The New Science of Cause and Effect [M]. Basic Books, 2018.
- [53] MILLETARI F, NAVAB N, AHMADI S A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation [J]. arXiv:1606.04797, 2016.
- [54] ZHAO Y, WANG X, CHE T, et al. Multi-task deep learning for medical image computing and analysis: A review [J]. Computers in Biology and Medicine, 2022, 153: 106496.
- [55] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces [J]. arXiv:2312.00752, 2023.
- [56] SHI Q, DUAN W, CHEN W, et al. PGP: Prior-Guided Pretraining for Small-sample Esophageal Cancer Segmentation [C] // 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2024: 3701-3704.
- [57] ZHOU J, CUI G, ZHANG Z, et al. Graph Neural Networks: A Review of Methods and Applications [J]. arXiv:1812.08434, 2018.



**WANG Yongquan**, born in 2001, post-graduate. Her main research interests include intelligent science technology and computer vision.



**SHI Qinglei**, born in 1982, Ph.D, senior scientist. His main research interests include artificial intelligence technology in the diagnosis and treatment of tumors and neurological diseases, establishment of an intelligent assisted diagnosis and

treatment decision-making system.

(责任编辑:何杨)