

深度学习在现代医疗领域中的应用^①

姚琼¹, 王觅也^{1,2}, 师庆科^{1,2}, 张梦娇¹, 邓悟¹

¹(四川大学华西医院 信息中心, 成都 610041)

²(四川大学华西医院 医疗信息化技术教育部工程研究中心, 成都 610041)

通信作者: 师庆科, E-mail: shiqingke@wchscu.cn



摘要: 深度学习作为人工智能领域最为活跃的研究分支, 近年来在计算机视觉、自然语言处理、语音识别等领域取得丰硕成果. 同时, 深度学习在医疗领域中的应用也逐渐成为研究热点, 并且在医学图像和信号处理、计算机辅助检测与诊断、临床决策支持、医疗信息挖掘和检索等方面取得了一些成功, 展现出了极大的应用前景. 本文在介绍深度学习原理和常用深度神经网络的同时, 结合相关文献和应用实践, 对深度学习在医疗系统中的应用场景和研究进展做了全面系统性地介绍. 同时, 本文还探讨了深度学习在现代医疗领域实施的难点和挑战, 并针对性地给出了一些解决方案或解决思路.

关键词: 深度学习; 人工智能; 医学信号和图像处理; 自然语言处理; 医院信息化

引用格式: 姚琼, 王觅也, 师庆科, 张梦娇, 邓悟. 深度学习在现代医疗领域中的应用. 计算机系统应用, 2022, 31(4): 33-46. <http://www.c-s-a.org.cn/1003-3254/8411.html>

Application of Deep Learning in Modern Healthcare

YAO Qiong¹, WANG Mi-Ye^{1,2}, SHI Qing-Ke^{1,2}, ZHANG Meng-Jiao¹, DENG Wu¹

¹(Information Center, West China Hospital of Sichuan University, Chengdu 610041, China)

²(Engineering Research Center of Medical Information Technology (Ministry of Education), West China Hospital of Sichuan University, Chengdu 610041, China)

Abstract: As the most active research branch of artificial intelligence, deep learning has achieved significant improvements in fields such as computer vision, natural language processing, and speech recognition, and its applications in healthcare have also become popular in recent years. Deep learning has made achievements in medical image and signal processing, computer-aided detection and diagnosis, clinical decision support, medical information mining and retrieval, etc., which shows great application prospects. While introducing the principles of deep learning and common deep neural networks, this study gives a systematic and comprehensive introduction to the typical applications and latest research progress of deep learning based on clinical practice and published papers. Moreover, the discussion of difficulties and challenges of deep learning in modern healthcare is included in this paper, and some solutions or ideas are also provided correspondingly.

Key words: deep learning; artificial intelligence; medical signal and image processing; natural language processing (NLP); hospital information

1 引言

当前, 以机器学习为代表的人工智能技术已经在

越来越多的行业领域中被广泛应用. 深度学习^[1-3]作为机器学习的重要组成部分, 近年来发展迅速, 并且在计

① 基金项目: 四川省科技计划 (2020YFS0092, 2020YFS0298)

收稿时间: 2021-06-18; 修改时间: 2021-07-14; 采用时间: 2021-07-27; csa 在线出版时间: 2022-03-22

算机视觉、自然语言处理、语音处理等领域取得了突破性的进展,甚至在某些场景上已经超过了人类专家的水平。

医疗领域相较于社交媒体、电子商务等行业来说更为的传统保守,同时也与广大群众的生命健康密切相关,现代医疗技术的发展可以保障并显著提升大众生活质量。自2018年开始,在医学和医疗卫生领域中与深度学习相关的研究热度急剧升温,越来越多的学者开始关注并尝试将深度学习相关技术应用在医疗领域中^[4],期待在解决疾病筛查、健康管理、诊治过程中的质量控制、优化资源配置等诸多问题中提供更有效的解决方式。但是,医疗领域由于其行业特殊性,有着专业性强、错误代价高、应用场景复杂、数据高度敏感等特点,虽然目前已经出现不少研究成果,但是要将这些技术落实到临床应用场景,还有很多问题亟待解决。

目前基于深度学习解决某些特定医疗问题的论文较多,但系统性地介绍深度学习在医疗领域应用的论述较少,虽然最新研究成果展现出来的评价指标不断地提高,但论及深度学习在医疗领域研究和实施的困难挑战却鲜有提及。本文尝试介绍深度学习的基本理论和常见深度神经网络模型,同时对深度学习在医疗领域中的应用场景进行梳理,介绍典型使用案例及研究进展。最后,本文还将对深度学习在实践应用中的常见问题,以及在医疗领域临床应用所面临的挑战进行介绍,并且结合行业的研究成果和作者的相关经验,给出一些可能的解决思路和方案。

2 深度学习

2.1 深度学习和机器学习

机器学习一直都是人工智能领域中的重要研究方向。其中,传统机器学习会将特征抽取和预测过程分开,需要领域专家和机器学习工程师协作完成数据处理、特征提取、特征转换等操作,然后设计合适的预测函数完成学习任务。传统机器学习的主要问题是模型的表达能力有限,需要花费大量精力从事特征工程,而且相关工作也基本凭借经验,或者通过大量的实验才能得到较好的效果。深度学习可以让原始数据在模型中经过多次转换,每次转换后可以形成更有效的特征表示,这种直接向模型提供原始数据,弱化甚至丢弃特征工程,以模型输出作为结果直接优化目标任务,称之为端到端的学习任务。

当前深度学习主要采用人工神经网络 (artificial neural network, ANN) 来实现,它是一种受生物神经系统工作方式启发而构造出的数学模型。人工神经网络由人工神经元及它们之间的连接构成,人工神经元的工作逻辑可以使用多种线性、非线性数学函数来定义,而它们的作用参数通过训练数据学习优化而来。实践证明,构造复杂多层的人工神经网络是当下深度学习最有效的实现方式。

2.2 常见深度神经网络模型

多层感知机 (multilayer perceptron, MLP) 是最简单的深度神经网络模型,当前在很多场景仍然被广泛使用。同时,卷积神经网络 (convolutional neural network, CNN)、循环神经网络 (recurrent neural network, RNN) 等理论先后被提出,而且近年在解决计算机视觉、语音处理、自然语言处理等领域获得巨大的成功^[2]。目前,应用场景中大量无标签数据集相对容易获取,因此深度自编码器 (deep autoencoder, DAE)^[5] 的使用频率也非常高。当前深度学习领域中,常见的神经网络模型结构如图1所示。

2.2.1 卷积神经网络

CNN 是近年来深度学习在计算机视觉领域取得突破性成果的基石,其网络结构主要由卷积层和池化层组成。图像处理领域中,卷积是常用的特征提取手段,不同卷积核可以得到不同特征图,而 CNN 的卷积层便是用来提取视野中的局部特征。卷积核是模型的参数,通过训练可以自动学习得到有效的特征表达,使同一个卷积核与图像所有像素做运算,可以避免参数膨胀。池化层主要缓解卷积层对位置过度敏感的问题,也降低特征的维度和训练参数的规模,常见的池化类型有最大池化层、平均池化层等。

2.2.2 循环神经网络

很多应用场景中,网络的输出除了与当前输出有关,还同之前一段时间的输出相关。RNN 便是一类具有短期记忆能力的神经网络,神经元在接受其它神经元信息的同时,也可以获取自身存储的状态信息,在自然语言、时序信号的处理中 RNN 被广泛使用,诸如机器翻译、文本生成、自动图像描述等都是比较典型的研究成果。理论上,RNN 可以处理任意长度的序列信息,但是 RNN 学习过程中如果记忆依赖的步程太长,容易产生梯度爆炸或者梯度消失的问题,所以现实情况下 RNN 只能实现“短期记忆”。通过引入门控机制

可以解决这类问题,目前广泛应用的包括门控循环单元(gated recurrent unit, GRU)和长短期记忆网络(long short-term memory, LSTM)^[2], LSTM通过引入输入门、遗忘门、输出门来控制信息的传递路径,使LSTM可以记录很长的步骤,增强了RNN类网络的实用性。

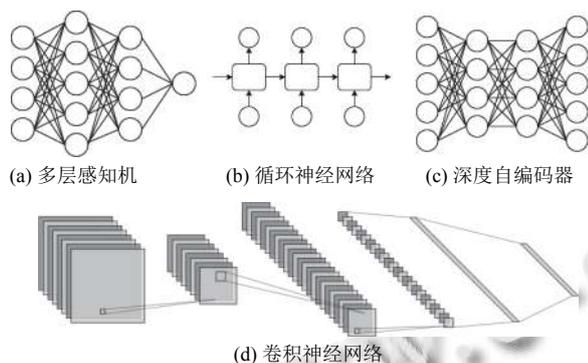


图1 常见深度神经网络模型结构

2.2.3 深度自编码器

DAE常常用来处理无标签数据集,由encoder和decoder两个部分组成,encoder用于将输入数据压缩到一个更小维度的表示,而decoder尝试从低维表示中重建原始数据,通过训练不断减少重建误差,使得神经网络能够从原始数据中提取最重要的部分。DAE的这种特性常用来抑制原始数据中的噪声,被应用在图像重建、消噪等领域^[6]。同时,最中间的隐藏层可以作为原始数据的嵌入表示,具有数据降维压缩的功能,与PCA算法相比,DAE通过多层网络和激活函数的作用,可以

对原始数据做更强的非线性变换表达^[5,7,8]。

3 深度学习在现代医疗领域中的应用

3.1 现代医疗领域的模型

现代医疗领域中常见数据类型大致可分为:结构化数据、影像视频类、文本类、时序生物信号类。

结构化数据通常以电子表格、数据库表等形式存储,比如患者个人基本信息、入院信息、检验结果等,因为其结构定义良好,在机器学习中经常使用。影像数据在医疗中也很常见,比如X光片、超声图像、MRI等。自由文本类数据可见于医生的诊断报告、医嘱信息、病历文书等。时序生物信号来自于仪器仪表的测量,例如病人的心电信号、脑电信号,以及各种动态生命监控数据等,在ICU中大量生理指标的实时检测和监控更加常见。电子病历(electronic medical records, EMRs)是医院信息化建设过程中产生的,管理着病人几乎所有医疗信息,而其中囊括的数据类型也是多种多样的。

3.2 医学影像相关业务

计算机视觉是深度学习最为成功的应用领域,同时医学影像也是医生用于诊断和评估疾病的常用手段,近年来深度学习处理医学影像的研究占据了深度学习在医疗领域应用中的绝大部分比重^[4]。在医疗领域中,通常需要处理的影像类型包括MRI、X光片、CT、超声、PET、组织切片病理图等,以及内窥镜、胶囊内视镜等视频数据,处理任务涵盖了图像分类、目标识别、图像分割、图像检索等,表1中对这些应用情形及学习任务的最终效果进行了总结^[9-26]。

表1 深度学习在医学影像相关任务的应用

任务	神经网络	数据模式	医学任务	结果	引用文献
图像分类	CNN (Inception v3)	皮肤彩色摄影图像	皮肤癌分类任务	ACC=0.721 (三分类), ACC=0.554 (九分类)	Esteva等 ^[10]
	CNN (ResNet18)	乳腺X-ray	乳腺癌5年内风险预估	AUC=0.70	Yala等 ^[11]
	CSDCNN	乳腺组织病理学图像	乳腺癌多分类预估	ACC=0.932	Han等 ^[12]
目标识别	CNN (ResNet)	胸部X-ray	肺部疾病的多分类预估和病患定位	AUC=0.863	Li等 ^[17]
	CNN (ResNet50)	结肠镜检查实时影像	息肉的实时识别和定位	ACC=0.964, AUC=0.991, SPEED=100 fps	Urban等 ^[18]
	CNN (SSD)	无线胶囊内窥镜检查影像	糜烂和溃疡的实时识别	AUC=0.958, SPEED=23 fps	Aoki等 ^[19]
图像分割	DAE	—	医学影像分割的后处理	—	Larrazabal等 ^[8]
	3D DSN	CT	三维肝脏分割	VOE=5.37%, VD=1.32%	Dou等 ^[22]
	CNN+VAE	MRI	脑部肿瘤分割	Dice=0.8219, Hausdorff=4.8289 mm	Myronenko ^[23]
图像检索	CNN	胸部X-ray	胸部病理图像检索	—	Anavi等 ^[24]
	CNN	Multi-Model	多模态图像检索	mAP=0.69	Qayyum等 ^[25]
	CNN (VGG19)	脑部MRI	脑肿瘤图像检索	mAP=0.961	Swati等 ^[26]

3.2.1 图像分类

图像分类任务是模型根据输入的图像进行预估,并输出一个对应的判别标签及其置信度.经典网络 AlexNet 于 2012 年被提出,当年以很大优势赢得了 ImageNet 大规模视觉识别挑战赛,从此基于 CNN 的应用开始受到了大家重视,并且开启了计算机视觉研究的新局势.此后,GoogLeNet、VGG、ResNet、DenseNet 等现代卷积神经网络也相继产生,在影像分类、分割等任务^[9]中被广泛的使用.

Esteva 等^[10]基于 Inception v3 主干网络,直接使用多达 13 万份带标注的临床影像数据来训练,带标注的训练数据集样本如图 2 所示,训练任务是检验该深度神经网络对于皮肤癌分类预估的性能,对照组是由 21 名皮肤科医生独立标注结果,结果显示深度学习的分类结果相比人类专家更好.同时,该实验证实了在大规模、高质量的标注数据集上,普通的 CNN 也能够产生很好的预估效果.

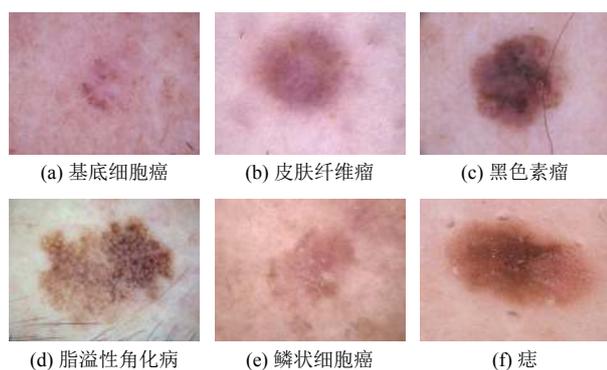


图 2 皮肤癌分类的训练样本样例

Yala 等^[11]结合就诊者的电子病历信息和乳腺造影图像,探究这些就诊者 5 年内患乳腺癌的风险,并且实际跟踪这些人在之后 5 年中是否实际患有乳腺癌作为标注,用来训练模型并评价算法.作者使用 Tyrer-Cuzick 模型作为基线,对比了逻辑回归、单独 ResNet 预估、以及结合两者混合模型的预估结果,最终发现结合电子病历信息和造影图像模型的 AUC 指标提升明显,这同时也启示结合领域内其它信息更有助于提升深度学习的预估和分类效果. Han 等^[12]基于乳腺组织病理学图像进行多分类实验,完成了对包含导管癌、纤维腺瘤、小叶癌等 8 类乳腺癌的分类任务,相对于传统二分类检测任务提供更丰富的临床诊断信息.

作者采用了一种结构化的深度卷积神经网络,优化了特征空间中相同分类和不同分类样本的相似度计算方法,在大规模数据集上,多分类学习任务的测试精度达到了 93.2%.

3.2.2 目标识别

目标识别任务用于对图像中特定结构或模式进行检测,并标示出对应目标的位置信息.在深度学习中,目标识别任务一般被分为两阶和一阶目标检测,其中两阶段目标识别算法以 R-CNN^[13],以及改进的 Fast R-CNN^[14]和 Faster R-CNN^[15]比较常见,而一阶目标识别算法最具有代表性的是 YOLO^[16].

R-CNN 是最早将 CNN 应用在目标检测任务中的算法,奠定了两阶目标识别的处理流程和算法框架,其实现步骤为:首先采用 Selective Search 方法生成若干候选区域,接着对每个候选区域使用 CNN 提取得到固定长度的特征向量,依次将特征向量输入到 SVM 进行分类判别,最终再使用一个线性回归器对得到的目标框进行精修. Fast R-CNN 的改进实现中,作者对整个图像提取特征图,降低对每个候选区域单独抽取特征带来的计算复杂度,并且将 SVM 和线性回归器集成到了深度神经网络中,避免数据交换的开销,整个系统架构上也更为优雅. Faster R-CNN 使用 RPN (region proposal network) 网络作为目标候选算法,利用 Anchor 机制将区域生成与卷积网络联系到了一起,提升了目标识别任务的效率和精确,使得两阶目标检测走向了实时化. YOLO 是最早提出的基于深度学习实现的一阶目标识别算法,该算法将原始图像分割成若干个网格,并对每个网格直接做前景、背景分类的判别预估,YOLO 算法的目标检测精度和 R-CNN 类算法大致相当,但是运行速度却快得多,性能上满足实时视频流处理,最新优化版本已经可以在移动设备上实时处理视频数据流了.

Li 等^[17]在较少的 X 光影像标注数据集上,用统一模型架构同时完成了肺部疾病的定位和判别任务,输出可能的病灶位置及对应疾病的种类和置信度.作者把输入图像划分成若干小片,再将这些小片合并成一个包,结合图像级别的标注信息执行多示例学习 (multi instance learning, MIL) 训练任务,使用多个独立判别器对多种胸部疾病在共享特征的情况下做独立训练和预估,损失函数综合这些判别器的误差.该方法提供了疾病多分类判别的实现思路,同时图像级别的弱标签标注可以有效降低训练数据的获取成本,在实践中更具

应用价值.除了2D静态医学影像,当前越来越多的应用场景在于完成视频流中特定目标的识别和跟踪任务,对目标识别的性能有了更高的要求. Urban等^[18]通过CNN完成对结肠镜检查中息肉的识别标注,其准确率达到96.4%,并且处理速度达到98 fps,完全胜任结肠镜检查的实时视频流处理.作者使用20段结肠镜检查视频,累计视频时长达到了5小时,让4名专业结肠镜医师在有和没有CNN辅助的情况下进行息肉的检查,结果前者检出的息肉组织数量比后者多了一倍,对于辅助结肠镜医生减少临床息肉漏检率意义重大,因为漏检会让病人错过最佳治疗时机,图3所展示的是在结肠镜检查过程中,算法自动对可疑组织的标注提示. Aoki等^[19]首次基于CNN完成对无线胶囊内窥镜检查中糜烂和溃疡病灶的检测和概率预估,检出准确度达到了88.2%,而且临床应用时适当提升模型的灵敏度,将更加有助于医师降低异常情况的漏检率.



图3 结肠镜检查中息肉组织的识别和标注提示

3.2.3 图像分割

图像分割任务相比目标识别更为精细,需要识别出对应目标并将其边界精确的描绘出来,该任务的输出一般是像素级别的分割描述,医学影像的语义分割容易受到周围组织的干扰,分割任务的难度大.图4是对胸部X光片的肺组织,以及对眼底图像的视网膜血管的分割效果图.

深度学习领域最早研究分割任务的是全卷积网络(fully convolutional network, FCN)^[20],FCN得名源自其网络结构将传统CNN最后若干个全连接层和Softmax操作以卷积层替代,于是整个网络都是由卷积层和池化层组成.不断地对图像执行卷积和池化操作,得到越来越小、越来越抽象语义化的特征描述,最末端的输出可以认为是每个像素作为目标分割的概率.虽然最终输出层的结果语义正确,但缺少细节信息,因此作者选择一些中间卷积层的结果做类似反卷积操作并且融合起来,补充了浅层网络提取的细节信息,这种融合多尺度特征的方法保证了最终分割结果的鲁棒性和精确

性. U-Net^[21]借鉴了FCN的思想,设计出更优雅的图像分割框架:它由contracting路径和expansive路径组成,前者不断地卷积和池化完成下采样操作,而后者拼接对等分辨率的特征后,执行反卷积完成上采样操作,不断恢复图像原始分辨率.整个网络框架形似一个完美对称U字型,而且每个分辨率层次的特征都保留并得到了充分的利用,实现了更丰富细致的分割结果.

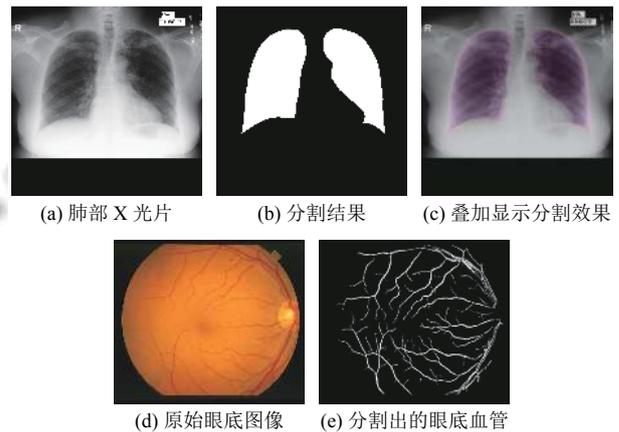


图4 肺部X光片和眼底图像的分割效果

Larrazabal等^[8]提出了一种基于DAE的图像分割后处理方案,该方案独立于图像分割流程,可以用于任何图像分割算法的后处理.在训练DAE的时候,随机对标注图像进行降级和扰动,比如切除某些区域、添加随机噪声等,以提升DAE的消噪能力,实践结果表明不仅对于随机森林这类传统分割算法,甚至对于未收敛的U-Net模型所得到的不完全分割结果,使用该方法处理后也有很好的最终分割效果.随着CT、MRI等越来越普及,对于3D图像的分割需求也越来越迫切. Dou等^[22]设计实现了3D深度监督网络(3D DSN)模型,用于对CT图像中的肝脏器官进行准确分割,相比传统3D CNN在体积重叠误差、相对体积差异等指标上都有明显改善,而且对比其它算法处理速率提升很多.作者通过对CNN的初始层、中间层提取出来的特征进行反卷积,然后独立运行预估结果并对比标注结果计算误差,在充分利用多尺度特征的同时也解决了梯度消失的难题,在少量标注训练样本的条件下实现了模型的快速收敛. Myronenko^[23]使用类似于U-Net网络的下采样-上采样的框架实现多模态3D MRI的脑组织的语义分割任务,作者在尝试各种网络设计和优化手段后,发现对训练图像执行尽可能大尺度的切片,

以及添加额外变分自动编码器 (variational autoencoder, VAE) 来辅助正则化可以显著提升分割精度, 该算法采用高性能 GPU 完成模型的训练任务, 获得了 BraTS 2018 比赛的冠军. 3D 图像的处理对于计算能力的要求极高, 但是这些年随着计算机硬件的发展和云计算的普及, 为这一类深度学习的应用提供了有力的支撑.

3.2.4 图像检索

医疗领域中, 医生不仅需要关注个体病患不同阶段、不同类型的医学影像, 还需要在医学影像库中快速检索相似部位和相似疾病的其它影像, 乃至多模态图像的跨库检索需求, 可以为医生提供更多的信息, 以便做病情诊断评估和临床决策. 伴随现代医学影像在临床中的大量使用, 医院的影像规模十分庞大, 快速、准确地提供图像检索技术成为了挑战.

医学图像检索的关键要素是特征表示和相似度描述. 传统上, 基于内容的图像检索 (content based image retrieval, CBIR) 的实现, 都是离线对入库图像完成图形的特征抽取, 比如常见的 SIFT、SURF 等特征描述, 然后将这些特征存储在数据库中, 在线检索的时候, 对检索图像做相同的特征抽取操作, 再采用欧式距离、余弦相似度等指标来衡量相似度, 和检索库中各个图像作相似度对比, 排序后返回检索结果.

上述传统实现方式依赖于影像专家和数据专家的经验来提取有效特征, 对于多模态影像的特征抽取和相似度比较也很困难, 而深度学习作为一种表示学习, 可以自动获得更有效的特征表示. Anavi 等^[24] 对于胸部 X 光片, 将传统图像处理技术锁提取的 SIFT-BoVW、LBP、Binary 特征同使用 CNN 提取的特征做对比, 结果显示使用 CNN 提取得到的特征相比传统图像特征检索效果最好. Qayyum 等^[25] 使用网络公开的多模态医学影像数据集, 对这些影像按照器官进行 24 个分类的粗粒度标注, 然后采用监督学习的方式训练 CNN, 并以网络最末端的 3 个全链接层输出作为特征表示, 测试得到多模态图像检索精度达到 69%, 相同条件下优于其它算法. Swati 等^[26] 利用深度学习检索 MRI 图像, 作者使用 VGG19 网络结构, 并加载使用已经在 ImageNet 中超过 120 万张带有 1000 分类标签的自然图像数据集上预先训练的模型参数. 从预训练模型开始, 将原始 VGG19 网络的 FC8 从 1000 分类调整为目标任务的 3 分类问题, 在已标注的 3000 多张 MRI 数据集上做模型精调, mAP 达到了 96%, 该流程是典

型迁移学习应用场景, 效果提升明显. Oliveira 等^[27] 使用领域索引 (domain index) 处理多库多模态医学影像, 设计出了可以弹性扩展的系统架构, 对于大型医院、医联体实现海量医学影像的快速检索具有一定的参考价值, 同时该框架还简化了机构对于多库医学影像的维护管理工作.

3.3 时序信号处理

传统处理时序类信号时, 一般需要使用傅立叶变换、小波变换等方式对原始信号进行预处理, 可以抑制信号噪声, 同时在频域处理信号往往更加简单有效. 然后, 领域专家凭借经验手动提取信号特征后, 再实现具体的分类、判别等任务. 深度学习在绝大多数场景都能够完成端到端的学习任务, 在使用深度学习处理时序信号时, 可以直接将原始信号输入到神经网络中训练模型优化目标任务.

Acharya 等^[28] 实现了对 ECG 信号心率失常疾病中 4 种类型进行分类, 他们构建了 11 层 CNN, 然后把 ECG 原始信号采样后直接送入深度神经网络执行训练任务, 在验证集上测试分类准确度达到了 92.5%, 而且作者增加了卷积层神经元个数后再次实验, 分类准确度提升到了 94.9%, 说明更多神经元构成的神经网络有助于在训练过程中获得更有效的特征. Hannun 等^[29] 也采用 CNN, 并将网络深度扩展到 34 层, 同时采用更大规模的标注数据集进行训练, 完成对心律失常疾病的 14 分类任务. 作者将算法的分类结果同几位心脏病专家综合后的标注结果作对比, 结果显示在绝大多数种类的分类任务中, 深度学习算法的 $F1$ 指标都显著高于专家手动标注结果. 还有 Rajput 等^[30] 的研究, 他们将时域一维 ECG 信号进行剪切分段后, 采用短时傅立叶变换和小波变换得到二维信号频谱图像, 然后采用图像处理的思路将它们送入到 DenseNet 中, 在相同测试集上效果要优于 Hannun^[29] 的实现. CNN 的特征学习优势明显, 同时由于 ECG 是时序信号, RNN 或 LSTM 可以挖掘时序上的依赖信息, 于是 Andersen 等^[31] 提出了 CNN-LSTM 混合深度网络结构, 首先利用 CNN 进行特征表示学习, 再利用 LSTM 进行序列学习, 也展示出了比较好的分类效果.

在临床条件下, ECG 采集过程很容易受到干扰, 对 ECG 进行噪声消除十分重要. 传统方式经常使用小波变换来消除噪声, 但实践中可能会因为阈值选取不当等问题导致信号重建产生偏差, 甚至会影响到最终

诊断结果的准确性. Chiang 等^[6]在 DAE 的基础上采用了全卷积网络代替原先的全连接网络, 处理结果的信噪比、均方误差等指标都明显好于单独使用 DNN、CNN 的处理结果.

3.4 自然语言处理

医疗领域中, 因为结构化的医疗数据格式规则、存取方便, 对其研究和应用挖掘地比较充分. 然而, 医疗领域中非结构化的自由文本数据更多, 除了医学教学和科研论文之外, 诸如临床诊断报告、影像检查报告、医嘱信息等内容, 都是医生认真检查、深思熟虑后的结论, 因此蕴含着极大的研究和应用价值. 同时,

近年来以医学知识图谱和电子病历为基础, 以人工智能推理为核心所构建的临床决策支持系统 (clinical decision support system, CDSS), 已经成为新一代医院信息系统建设的重要内容, 对于提高临床医护人员诊断和护理的工作效率, 提高医疗质控水平, 减少医疗事故的发生意义重大, 图 5 展示了典型的临床决策支持系统的典型系统架构. 但是, 自然语言处理一直都是机器学习领域的研究难点, 而且医疗领域的文本数据还涉及到大量的领域内部命名实体、行业术语及表述习惯等问题, 因此基于深度学习的自然语言处理在医疗领域的应用相比而言成熟度较低.

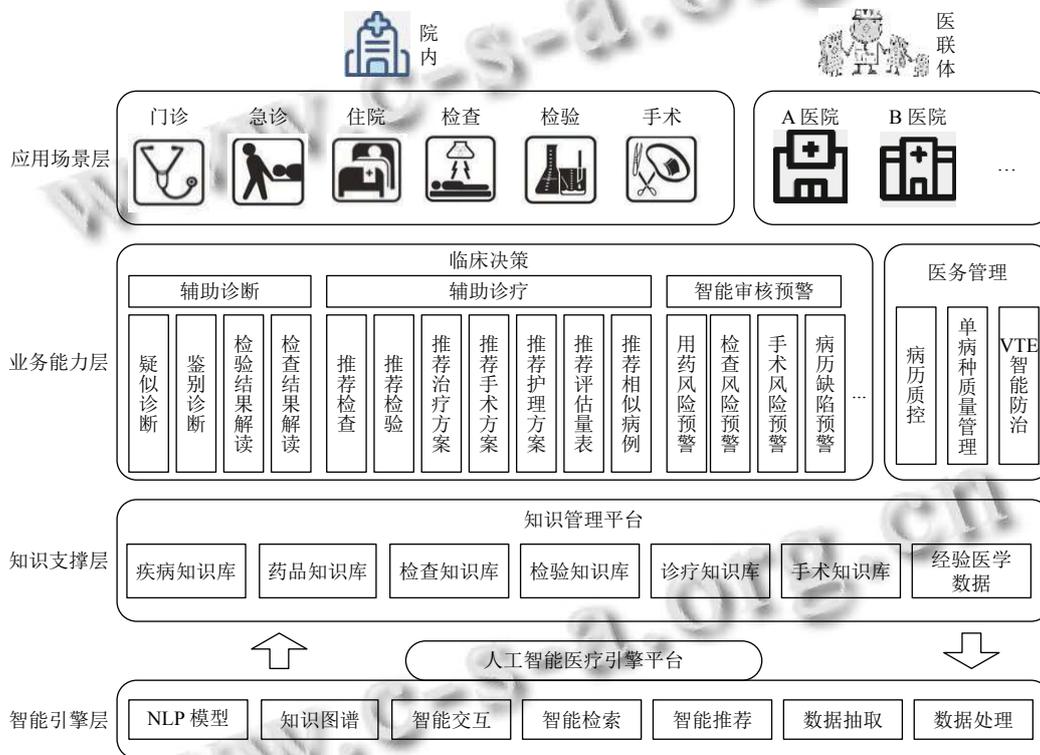


图 5 临床决策支持系统架构图

3.4.1 疾病分类编码

ICD (international classification of diseases) 是世界卫生组织制定和维护的人类疾病标准化、统一化的分类系统, 各个国家和地区对本地疾病分类管理也有相关规范, ICD 对于医疗卫生资源分配、公共卫生建设意义重大^[32]. 目前, 疾病分类编码仍旧采用人工标注的方式来完成, 但是由于疾病种类数目太过繁多, 很多疾病类间差异也很细微、界定模糊, 采用人工标注需要消耗大量精力, 而且难以保证标注结果的准确性和一

致性. 因此, 学者们一直尝试采用机器学习的手段完成该任务, 期望通过自然语言处理技术, 根据诊断病历自动完成相关疾病的编码任务. ICD 分类多、疾病的分布不均衡, 是一个典型极端性多分类问题.

Baumel 等^[32]对比使用 SVM、CBOW、CNN 和 HA-GRU 四种算法, 采用公共的 MIMIC 数据集完成 ICD-9 疾病编码, 其中 SVM 采用了 TD-IDF 对文本单词计算权重, 而 CBOW 和 CNN 都使用了稠密的嵌入向量表示字符. HA-GRU 是一种层级的双向 GRU 网

络,采用底层双向 GRU 网络编码语句,而级联的上层双向 GRU 网络用于接收底层所有语句编码,这种设计是考虑到 GRU 处理太长文本会有性能瓶颈和效果损失,结果显示 HA-GRU 相比基线算法效果提升明显. Qiu 等^[33]采用人工标注的 942 份分类结果作为训练数据集实现对 ICD-O-3 癌症疾病编码,在使用 TD-IDF 作为单词权重条件下,结合朴素贝叶斯、逻辑回归、SVM 等传统分类器,同 CNN 实现的自动特征提取和分类的结果做对比,结果使用 CNN 方式分类优势明显. Mullenbach 等^[34]采用了带 Attention 机制的 CNN,使用 MIMIC-III 数据集做 ICD-9 编码,测试效果相比之前的 HA-GRU 算法^[32]也有明显的提升.

3.4.2 文本数据挖掘

文本数据挖掘可以应用于电子病历结构化、临床决策支持、异常事件检测、信息语义化检索等场景.传统上对自由文本信息绝大多数采用基于规则的专家系统处理,但是后期的维护工作极其繁重复杂,因此越来越多基于机器学习的方法被尝试用于解决这类难题.目前,通过对自由文本形式的资料进行清洗和整理,识别命名实体,推断实体间的关系,通过语义分析建立强大的医学知识图谱和知识库,在疾病风险评估、智能辅助诊疗、医疗质量控制及医疗知识问答等智慧医疗领域都有着很好的发展前景.

Yala 等^[35]采用 Boost 算法处理乳腺病理学报告,完成多达 20 种有关疾病类型、器官组织特性等数据的自动提取分类,并将这些数据整理后进行结构化存储和展示.其总体分类准确度超过了 90%,基本同基于规则的处理方法效果相当,但是极大节约了工作量. Borjali 等^[36]探索了采用深度学习从纯文本的医疗报告中提取相关信息,完成对医疗不良事件的检测统计工作,并以髌关节置换术后脱位的案例进行试验.作者设计实现了类似于 HA-GRU^[32]的层级结构 BiLSTM 和 CNN 两种深度网络模型,最终结果显示基于深度学习方案的 Kappa、分类精度指标明显高于 K-NN、随机森林、SVM 等为代表的传统机器学习算法,并且 CNN 的效果要好于层级 BiLSTM 的实现.值得注意的是,IBM、Google、Amazon、腾讯等^[37,38]大型商业公司,以及国内外许多医疗科技企业,也都在积极布局深度学习在医疗领域的应用,尝试将他们在深度领域的经验技术赋能于医疗健康事业中.

3.4.3 诊断报告的自动生成

医护人员每天都要花费很多的时间处理文本工作^[7],虽然报告模版、常用语填充等方式简化这方面的工作,但是这类工具既不灵活也不智能,仍然耗费医护人员宝贵的时间精力.当自动图像描述的功能被提出后,自动生成医学影像报告的功能也开始被大量研究了.自动图像描述的报告会基于对医学影像或信号的处理结果来生成,在提高医师的工作效率,同时提供独立的诊断结果供参考和对比.

Wu 等^[39]将自动图像描述应用在糖尿病视网膜眼底图像的 diagnosis 上,作者直接使用经典的 CNN-LSTM 网络结构,即首先使用 CNN 抽取特征,再依据特征使用 LSTM 生成对应的文字描述,作者的训练数据集使用了 370 张临床眼底图像,并对每张图像人工添加 5 种自然语言描述作为标注,其测试结果显示对病变眼底图像生成的报告准确度能达到 90%,但当测试集包含正常眼底图像后,整体准确度下降到只有 60% 左右,算是诊断报告的最简单尝试. Liu 等^[40]提出通常情况的自动图像描述所生成的内容都比较简短,同时研究重心也偏向于生成文本的可读性,但是临床报告常需要生成很长的描述段落,而且相较于语言的可读性,对结果描述的准确性是首要考量的.作者使用了 CNN-RNN-RNN 网络结构, CNN 用于学习图像特征,然后 RNN 依据图像特征生成话题,而后一个 RNN 再依据生成的话题和图像特征生成自然语言描述,最后再使用强化学习对生成的结果进行优化,最终生成报告的准确性和可读性的质量很高.

不过, Pino 等^[41]使用随机返回、持续正向返回、相似图像报告返回等简单验证手段,发现当前所谓领先算法所使用的评价指标并不可靠,比如 ROUGE、BLEU 等经常是用于机器翻译方面的应用,用于医学影像报告自动生成这类应用场景,上述评价指标或许不太合适,因此还需设计更加科学合理的评价指标.

4 深度学习在现代医疗领域中实施的挑战

4.1 深度学习的固有问题

深度学习领域中,通过大量带标注数据集进行训练,几乎都可以得到一个预估误差很小的神经网络,但是缺少相关理论和方法来解释其因果关系,这是一个长期困扰深度学习的难题.对于诸如医疗领域这种关

键场景,“黑盒”应用具有相当大的隐患,尤其对于涉及到责任、纠纷等问题,这种不能解释的系统很难得到医生的信任和接纳.目前,绝大多数的智能医疗解决方案至多充当为一个独立建议者的角色,最终结论仍然需要由医生和护士的确认.同时,无法解释性也就意味着无法针对性地对算法作出改良,更多情况下只能通过不断暴力尝试来“拼凑”出一个恰到好处的模型.

针对如何解决深度模型解释性问题,当前提出的CAM和Grad-CAM^[42-44]或许有所帮助,但仍属于比较粗糙的办法.原理上,CNN随着不断卷积和池化操作,越接近末端的网络学习到的特征越抽象,但也越接近于任务目标.因此,Grad-CAM选择CNN最末端的卷积层,然后计算流入该层神经元相对某个分类的权重信息,该权重信息表示了特征相对于这个分类的敏感程度,进而得到热力图,图6展示了图像分类和图像分割任务中的Grad-CAM热力图,颜色越深的部位则表示对最终计算结果支持程度越重要.目前很多基于深度学习的应用都会附上热力图以证明模型的可信度^[40,43,45],但这种方式难以应用在其它类型的深度神经网络类型,而且随着深度网络的结构越来越复杂,要彻底理解深度神经网络的因果关系还很难.Doppalapudi等^[46]在对肺癌生存期预测的实验中,深度学习同样显示出比传统机器学习更好的预估效果,作者使用SHAP(shapley additive explanations)来分析输入特征对最终预估结果的重要性,得到良性肿瘤数目、受检查者年龄、是否接收手术是决定生存期最为重要的因素,该结论同医生的主观经验相符.

对于已经构建完成的神经网络模型,在实验和应用场景中常会发生模型参数或网络结构的微调,都可能会导致模型性能有巨大的变化,深度学习的鲁棒性也有待研究.深度模型也易于遭遇对抗攻击、模型和数据投毒等安全问题,确保其安全稳定地运行仍面临着不小的挑战,目前也有不少学者开始专注于研究这类问题的解决方式^[47].

4.2 训练数据集

在机器学习领域,绝大多数应用场景都是基于监督学习完成,模型效果很大程度上取决于已标注训练数据集的规模和质量^[10,29],但是医疗领域的数据存在诸如标注数目有限、样本分布不均衡、字段数据缺失现象普遍等问题.

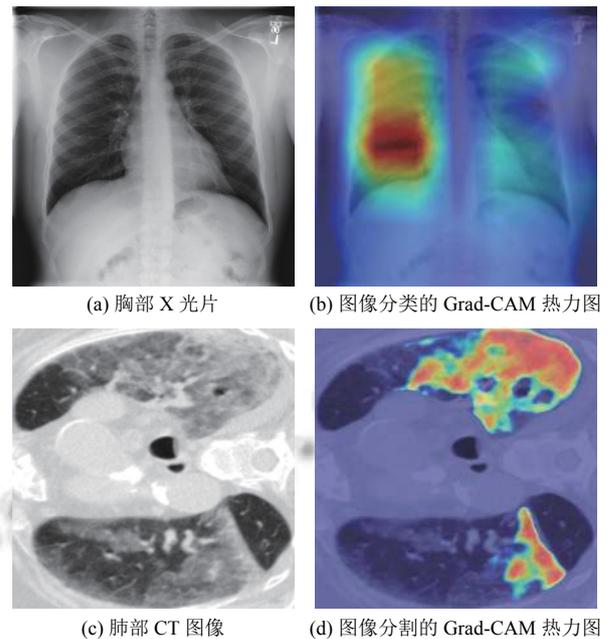


图6 图像分类和分割任务的Grad-CAM热力图

首先,医疗领域作为一个高度专业化领域,只有经验丰富的医师产生的标注结果才是相对可靠的,这决定了生成高质量标注数据集是一个产量低、代价高昂的任务.而且,标注结果本身也存在一定的主观性,即使都是经验丰富的医师,他们的标注结果常常也会很不一致^[48],而且同一个人对同样样本的标注也会经常有前后不一致的情况,使得模型的训练和评价任务更为艰难.医疗领域的样本分布也是极不均衡的,现实场景中大量样本都是正常或常见疾病,严重疾病都是罕见匮乏的,直接依赖这样的样本分布难以训练出有效可靠的模型.机器学习对于常见案例可以给出相对可靠的预估,而罕见或训练样本中没遇到的疾病,其预估结果往往难以预料^[3].同时,现实情况中数据缺失也很常见,医生会根据患者的个体情况作出不同的检查和诊疗方案,这类样本是否选择,以及相关字段如何补齐,预估结果是否可靠也是需要考虑的问题.

通过数据增强(data augmentation)技术可以扩充训练样本,经典图像增强手段包括更改图像对比度、叠加噪声,更改图像亮度、饱和度、对比度,以及对图像进行旋转、切割、缩放、形变等操作,然后对标注结果做必要的调整,可以有效扩充标注数据^[43].在训练U-Net网络的时候^[21],作者将有限的标注样本进行图形学弹性变换,不仅扩充和平衡了标注训练样本,同时也让模型“学习”了形变相关的知识,在医学影像中软体

组织的弹性形变是很常见的情形。Zhang 等^[49]研究了基于深度学习的 DST (deep stacked transformations) 神经网络完成图像增强扩充训练样本后, 验证模型迁移在未知测试数据集上的泛化能力, 结果显示 DST 神经网络的稳定性相比传统图像增强技术和基于 CycleGAN 的图像增强技术提升很多, Dice 指标提升超过 30%, 而且采用 DST 神经网络在少量标注数据上进行图像增强后训练模型, 然后在大规模未知数据集上的预估已经可以和最先进的监督学习在大规模标准样本上训练模型的预估效果相媲美了。Zhao 等^[50]发现 MRI 在患者脑灰质和脑白质的病变检测和分类中十分有效, 但是因为检测价格、设备限制等各项因素, 脑部 CT 影像比 MRI 影像更为常见, 但同时 CT 影像对于软体组织对比度很差, 很难直接用来进行脑灰质、脑白质、脑脊液的病变检测。作者通过改进 U-Net 网络, 使用带标注的 CT 影像构造出对应的 MRI 影像, 因为构造前后的图像是自然配准的, 所以 MRI 可以复用原标注信息执行监督学习, 测试结果证明了利用 CT 影像完成软组织标注和诊断任务的可行性。主动学习也可以用于解决标注数据缺乏问题, 算法的基本原理是系统塞选出不确定性样本交由专家标注, 并将标注结果放入训练集中持续优化模型。Kuo 等^[51]在使用基于不确定性样本选择机制时, 引入预估标注时间作为代价参考因素, 该模型综合标注收益和标注代价, 让医生在宝贵的时间中产生更多有价值的标注结果。

迁移学习^[26,52-54]目前在深度学习中经常被使用。训练初始可以使用已有的大量高质量标注的自然图像训练模型, 先“学习”线条、形状、边缘等知识, 最后再将该网络较前端部分的输出作为特征抽取算子, 或者固定其较浅部分的模型参数, 再根据目标任务优化末端部分的模型参数, 就可以使用较少标注数据使整个模型快速收敛^[26]。为了节省计算资源, 目前越来越多的场景采用标准主干网络并加载预训练模型, 再根据目标任务调优的迁移学习方案。迁移学习在 2D 图像的训练任务中十分常见, 但是在 3D 图像处理中研究案例很少, 相邻图像的上下文相关信息无法挖掘, 即使尝试使用 RNN 或 LSTM 来学习相邻图像的相关特征, 考虑到计算量和模型复杂度等因素几乎也是无法实现的, 而且通常 3D 图像标注数据很少, 因而模型很难训练。为此, Liu 等^[55]提出了 3D AH-Net 模型, 它使用 ResNet 作为主干网络, 并基于 ImageNet 的预训练模型作为起

点, 然后再将目标任务中 3 张相邻的 2D 图像模拟成 RGB 三个通道对编码器网络做精调, 在保持网络结构不变的同时引入相邻上下文关系。作者设计了全新解码器代替原先解码器, 把 3D 卷积操作拆分成 2D 卷积操作和 1D 卷积操作, 在充分利用单个图像内部特征的同时, 保持图像间相关性输出一致。最终, 在乳腺病变检测、肝脏分割两个测试任务中, 该网络效果领先传统网络 (例如 3D U-Net), 而且在训练和预估任务上也有着明显的性能优势。Liang 等^[56]研究了医学图像应用中深度学习的模型泛化问题, 实验对比了迁移学习中 3 种常用实现手段的泛化性能, 针对目标数据集的模型微调的泛化收益最为明显。

课程式学习^[57,58]作为一种训练策略, 值得对于医疗领域这类标注数据少、样本分布不均衡场景尝试。该训练策略启发自人类学习过程, 普通人在学习过程中常从较简单的任务开始, 接着不断增加学习难度, 这种难度逐步增加的学习过程更有效率。但是, 如何确定样本的难易程度也是比较微妙的事情, Jiménez-Sánchez 等^[58]的研究表明: 如果先学习困难的样本, 再学习比较容易的样本, 最终模型的预估效果反而会变差, 而且这种难易程度和人类直观感受的难易程度是不一致的。

4.3 模型的迁移泛化和实施

深度学习在医疗领域的应用研究, 研究成果部分是基于特定医疗机构的内部数据, 往往难以复现。而绝大多数都是基于网络公开数据集, 这类数据集规模普遍都很小, 而且一些字段信息也删除掉了。考虑到各医疗机构的就诊条件、设备差异、运行参数、病人分布等因素都可能存在差异性, 如此小规模训练集是否能够代表真实应用场景的样本分布, 以及训练所得的模型在其它设备、其它机构是否同样有效还存在巨大挑战^[47]。当前看来, 除了采用超大规模、多样性的标注数据训练模型外, 通过数据增强、多任务学习、迁移学习等手段对提高模型的泛化能力对增加临床使用的适应能力具有一定的意义^[43,49]。

深度学习结合医疗领域本来就是交叉课题, 通常需要数据专家、算法、工程技术、医疗领域专家的紧密合作才行。但目前情形而言, 医疗机构本身对这方面的研究精力和资源投入有限, 同外部机构合作研究交叉课题机会也不多, 整个流程还需要医院管理层和伦理委员会的审批和监督, 如此严格的高准入门槛阻碍了大部分研究机构和学者获取真实、有效、丰富的

医疗数据,着力共同攻克相关领域难题的机会。

传统医疗设备和信息系统,在实施部署后便进入常规维护流程,后期的维护工作都相对有限,而基于深度学习的医疗应用,后期维护任务也更加的复杂。深度学习需要持续跟踪使用场景,通过不断优化模型才能保持和提升效果,而且随着新数据的产生,或者医疗设备、使用应用场景发生变化,原有模型都需要不断地更新维护。现代医学影像的分辨率越来越高,同时3D影像技术应用越来越广泛,因此还需要购置先进的硬件设备以提供强大的计算能力。当前深度学习只擅长于对特定应用场景进行建模,而大型综合医疗机构中的科室数目、疾病类型、服务场景都是海量的,这意味着巨大应用潜力的同时,复用性低也意味着巨大的投入负担,深度学习的应用实施对基础信息设施有着较高的要求,这无疑会增加医疗机构运营和管理维护的成本。总之,对于医疗机构来说,深度学习应用是一个需要持续性、高投入的场景,对此医疗机构的态度势必会更加的谨慎。

4.4 隐私合规性问题

当前公众对于个人隐私问题越来越关注,尤其对于医疗健康信息尤为敏感,数据保护也是当前医院管理的重中之重。美国在1996年就签署颁布了健康保险流通与责任法案(health insurance portability and accountability act, HIPAA), HIPAA 包含了一系列规定来保障受保护健康信息(protected health information, PHI)的安全性和隐私性, PHI 涉及到年龄、联系方式、社会保险号等字段, HIPAA 适用于包括直接接触病人并处理病人数据的医院、医疗服务提供商、研究机构和保险公司等主体。欧盟于2018年也出台了通用数据保护条例(general data protection regulation, GDPR)用于个人敏感数据的保护。我国对健康隐私的保护机制也在不断地健全和完善。2021年1月1日实施的《中华人民共和国民法典》中定义了个人信息包括自然人的姓名、出生日期、身份证件号码、生物识别信息、住址、电话号码、电子邮箱、健康信息、行踪信息等,同时要求医疗机构及医务人员应当对患者的隐私和个人信息保密。

虽然 HL7、IHE、DICOM、openEHR 等标准的制定和实施,在技术层面上解决了医疗数据交换和流通的难题,但对于隐私问题的过分担忧会严重阻碍对这些信息的挖掘利用和互操作^[59],对医疗数据实现脱敏

是一种解决思路。相关研究^[60]已经尝试用机器学习和深度学习的方式,自动对患者个人医疗信息进行 PHI 归类判别,以代替传统的手动或基于规则的专家系统来完成相应工作。Catelli 等^[54]尝试将文本嵌入表示和 BiLSTM+CRF 模型应用于 COVID-19 意大利语电子病历数据集的去标识化操作,其评测效果优势显著,也展现出通过深度学习对医疗数据实现脱敏去标识化后,共享医疗信息快速应对公共卫生事件的重要价值和意义。相信终有一天,公众和社会对于患者健康数据的使用将会更加安全理性。

5 结束语

本文介绍了深度学习的原理和常见深度神经网络模型,结合实践介绍了深度学习在医疗领域中的影像和视频处理、信号处理、自然语言处理等典型医疗数据的应用,包含了科学研究和临床应用的最新研究进展、研究热点等问题,其中以医学影像的研究成果最为显著。本文还对深度学习在医疗领域中的应用困难和挑战进行了讨论,其中包括模型的解释性、稳定性等深度学习的固有难题,以及医疗领域标注数据缺乏、样本分布不均衡等行业特定问题,目前主流方法通过使用数据增强、迁移学习、课程学习等技术解决这类问题。同时,面对大量存在的无标注数据和弱标注数据,无监督学习、弱监督学习、半监督学习等算法也有很大的研究价值。深度学习在医疗领域的应用是一个交叉课题,需要多领域专家协作才能取得进展,虽然数据交换在技术上不再困难,但考虑到隐私、监管合规等因素,数据壁垒是目前发展的重要阻碍。

中国是一个幅员辽阔、人口众多,同时医疗资源分布很不均衡的国家,伴随着人口老龄化程度不断地加深,医疗服务的供需矛盾将更加的尖锐,而运用现代信息技术有助于解决医疗资源的分配和医学知识技能的共享问题^[10,56]。近年来深度学习在医学领域的应用有着丰富的收获,是一个极具意义的研究方向,但是距离临床应用还有很多问题亟待解决,期待深度学习技术能够更快更好地造福于人类健康事业。

参考文献

- 1 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521(7553): 436-444.
- 2 Goodfellow I, Bengio Y, Courville A. 深度学习. 赵申剑, 黎

- 戡君,符天凡,等译.北京:人民邮电出版社,2017.201–253.
- 3 Yuille AL, Liu CX. Deep nets: What have they ever done for vision? *International Journal of Computer Vision*, 2021, 129(3): 781–802.
 - 4 National Library of Medicine. <https://pubmed.ncbi.nlm.nih.gov/?term=deep+learning>. [2021-02-10].
 - 5 Gondara L. Medical image denoising using convolutional denoising autoencoders. 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW). Barcelona: IEEE, 2016. 241–246.
 - 6 Chiang HT, Hsieh YY, Fu SW, *et al.* Noise reduction in ECG signals using fully convolutional denoising autoencoders. *IEEE Access*, 2019, 7: 60806–60813.
 - 7 Esteva A, Robicquet A, Ramsundar B, *et al.* A guide to deep learning in healthcare. *Nature Medicine*, 2019, 25(1): 24–29.
 - 8 Larrazabal AJ, Martínez C, Glocker B, *et al.* Post-DAE: Anatomically plausible segmentation via post-processing with denoising autoencoders. *IEEE Transactions on Medical Imaging*, 2020, 39(12): 3813–3820.
 - 9 续宝红,丁冲.卷积神经网络在阿尔茨海默病诊断中的应用研究. *生物医学工程学杂志*, 2021, 38(1): 169–177, 184.
 - 10 Esteva A, Kuprel B, Novoa RA, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017, 542(7639): 115–118.
 - 11 Yala A, Lehman C, Schuster T, *et al.* A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology*, 2019, 292(1): 60–66.
 - 12 Han ZY, Wei BZ, Zheng YJ, *et al.* Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific Reports*, 2017, 7(1): 4172. [doi: [10.1038/s41598-017-04075-z](https://doi.org/10.1038/s41598-017-04075-z)]
 - 13 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Columbus: IEEE, 2014. 580–587.
 - 14 Girshick R. Fast R-CNN. *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago: IEEE, 2015. 1440–1448.
 - 15 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149.
 - 16 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2016. 779–788.
 - 17 Li Z, Wang C, Han M, *et al.* Thoracic disease identification and localization with limited supervision. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8290–8299.
 - 18 Urban G, Tripathi P, Alkayali T, *et al.* Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology*, 2018, 155(4): 1069–1078. e8.
 - 19 Aoki T, Yamada A, Aoyama K, *et al.* Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointestinal Endoscopy*, 2019, 89(2): 357–363. e2.
 - 20 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston: IEEE, 2015. 3431–3440.
 - 21 Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015. 234–241.
 - 22 Dou D, Yu LQ, Chen H, *et al.* 3D deeply supervised network for automated segmentation of volumetric medical images. *Medical Image Analysis*, 2017, 41: 40–54.
 - 23 Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. 4th International Conference on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Granada: Springer, 2018. 311–320.
 - 24 Anavi Y, Kogan I, Gelbart E, *et al.* A comparative study for chest radiograph image retrieval using binary texture and deep learning classification. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Milan: IEEE, 2015. 2940–2943.
 - 25 Qayyum A, Anwar SM, Awais M, *et al.* Medical image retrieval using deep convolutional neural network. *Neurocomputing*, 2017, 266: 8–20.
 - 26 Swati ZNK, Zhao QH, Kabir M, *et al.* Content-based brain tumor retrieval for mr images using transfer learning. *IEEE Access*, 2019, 7: 17809–17822.
 - 27 Oliveira PH, Scabora LC, Cazzolato MT, *et al.* Employing domain indexes to efficiently query medical data from multiple repositories. *IEEE Journal of Biomedical and Health Informatics*, 2019, 23(6): 2220–2229.
 - 28 Acharya UR, Fujita H, Lih OS, *et al.* Automated detection of arrhythmias using different intervals of tachycardia ECG

- segments with convolutional neural network. *Information Sciences*, 2017, 405: 81–90.
- 29 Hannun AY, Rajpurkar P, Haghpanahi M, *et al.* Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 2019, 25(1): 65–69.
- 30 Rajput KS, Wibowo S, Hao C, *et al.* On arrhythmia detection by deep learning and multidimensional representation. arXiv: 1904.00138v4, 2019.
- 31 Andersen RS, Peimankar A, Puthusserypady S. A deep learning approach for real-time detection of atrial fibrillation. *Expert Systems with Applications*, 2019, 115: 465–473.
- 32 Baumel T, Nassour-Kassis J, Cohen R, *et al.* Multi-label classification of patient notes a case study on ICD code assignment. arXiv: 1709.09587v3, 2017.
- 33 Qiu JX, Yoon HJ, Fearn PA, *et al.* Deep learning for automated extraction of primary sites from cancer pathology reports. *IEEE Journal of Biomedical and Health Informatics*, 2018, 22(1): 244–251.
- 34 Mullenbach J, Wiegrefe S, Duke J, *et al.* Explainable prediction of medical codes from clinical text. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. New Orleans: Association for Computational Linguistics, 2018. 1101–1111.
- 35 Yala A, Barzilay R, Salama L, *et al.* Using machine learning to parse breast pathology reports. *Breast Cancer Research and Treatment*, 2017, 161(2): 203–211.
- 36 Borjali A, Magnéli M, Shin D, *et al.* Natural language processing with deep learning for medical adverse event detection from free-text medical narratives: A case study of detecting total hip replacement dislocation. *Computers in Biology and Medicine*, 2021, 129: 104140. [doi: [10.1016/j.compbimed.2020.104140](https://doi.org/10.1016/j.compbimed.2020.104140)]
- 37 Mehta N, Devarakonda MV. Machine learning, natural language programming, and electronic health records: The next step in the artificial intelligence journey? *Journal of Allergy and Clinical Immunology*, 2018, 141(6): 2019–2021. e1.
- 38 Shang H, Sun ZQ, Yang W, *et al.* Leveraging other datasets for medical imaging classification: Evaluation of transfer, multi-task and semi-supervised learning. *22nd International Conference on Medical Image Computing and Computer Assisted Intervention*. Shenzhen: Springer, 2019. 431–439.
- 39 Wu LH, Wan C, Wu YQ, *et al.* Generative caption for diabetic retinopathy images. *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*. Shenzhen: IEEE, 2017. 515–519.
- 40 Liu GX, Hsu TMH, McDermott M, *et al.* Clinically accurate chest X-ray report generation. *Proceedings of the 4th Machine Learning for Healthcare Conference*. Ann Arbor: PMLR, 2019. 249–269.
- 41 Pino P, Parra D, Messina P, *et al.* Inspecting state of the art performance and NLP metrics in image-based medical report generation. arXiv: 2011.09257v2, 2020.
- 42 Selvaraju RR, Cogswell M, Das A, *et al.* Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 618–626.
- 43 Hu SP, Gao Y, Niu ZM, *et al.* Weakly supervised deep learning for COVID-19 infection detection and classification from CT images. *IEEE Access*, 2020, 8: 118869–118883.
- 44 Gotkowski K, Gonzalez C, Bucher A, *et al.* M3d-CAM: A PyTorch library to generate 3D data attention maps for medical deep learning. arXiv: 2007.00453, 2021.
- 45 Karakanis S, Leontidis G. Lightweight deep learning models for detecting COVID-19 from chest X-ray images. *Computers in Biology and Medicine*, 2021, 130: 104181. [doi: [10.1016/j.compbimed.2020.104181](https://doi.org/10.1016/j.compbimed.2020.104181)]
- 46 Doppalapudi S, Qiu RG, Badr Y. Lung cancer survival period prediction and understanding: Deep learning approaches. *International Journal of Medical Informatics*, 2021, 148: 104371. [doi: [10.1016/j.ijmedinf.2020.104371](https://doi.org/10.1016/j.ijmedinf.2020.104371)]
- 47 Qayyum A, Qadir J, Bilal M, *et al.* Secure and robust machine learning for healthcare: A survey. *IEEE Reviews in Biomedical Engineering*, 2021, 14: 156–180.
- 48 Joskowicz L, Cohen D, Caplan N, *et al.* Inter-observer variability of manual contour delineation of structures in CT. *European Radiology*, 2019, 29(3): 1391–1399.
- 49 Zhang L, Wang XS, Yang D, *et al.* When unseen domain generalization is unnecessary? Rethinking data augmentation. arXiv: 1906.03347v2, 2019.
- 50 Zhao C, Carass A, Lee J, *et al.* Whole brain segmentation and labeling from CT using synthetic MR images. *8th International Workshop on Machine Learning in Medical Imaging*. Quebec City: Springer, 2017. 291–298.
- 51 Kuo WC, Häne C, Yuh E, *et al.* Cost-sensitive active learning for intracranial hemorrhage detection. *21st International Conference on Medical Image Computing and Computer Assisted Intervention*. Granada: Springer, 2018. 715–723.
- 52 Yosinski J, Clune J, Bengio Y, *et al.* How transferable are

- features in deep neural networks? Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: ACM, 2014. 3320–3328.
- 53 陈道争, 江倩. 基于卷积神经网络和迁移学习的肺结节检测. 计算机工程与设计, 2021, 42(1): 240–247.
- 54 Catelli R, Gargiulo F, Casola V, *et al.* A novel COVID-19 data set and an effective deep learning approach for the de-identification of Italian medical records. IEEE Access, 2021, 9: 19097–19110.
- 55 Liu SQ, Xu DG, Zhou SK, *et al.* 3D anisotropic hybrid network: Transferring convolutional features from 2D images to 3D anisotropic volumes. 21st International Conference on Medical Image Computing and Computer Assisted Intervention. Granada: Springer, 2018. 851–858.
- 56 Liang X, Nguyen D, Jiang SB. Generalizability issues with deep learning models in medicine and their potential solutions: Illustrated with cone-beam computed tomography (CBCT) to computed tomography (CT) image conversion. Machine Learning: Science and Technology, 2021, 2(1): 015007.
- 57 Wei J, Suriawinata A, Ren B, *et al.* Learn like a pathologist: Curriculum learning by annotator agreement for histopathology image classification. arXiv: 2009.13698v1, 2020.
- 58 Jiménez-Sánchez A, Mateus D, Kirchoff S, *et al.* Medical-based deep curriculum learning for improved fracture classification. 22nd International Conference on Medical Image Computing and Computer Assisted Intervention. Shenzhen: Springer, 2019. 694–702.
- 59 Shrivastava U, Song JH, Han BT, *et al.* Do data security measures, privacy regulations, and communication standards impact the interoperability of patient health information? A cross-country investigation. International Journal of Medical Informatics, 2021, 148: 104401. [doi: [10.1016/j.ijmedinf.2021.104401](https://doi.org/10.1016/j.ijmedinf.2021.104401)]
- 60 Dernoncourt F, Lee JY, Uzuner O, *et al.* De-identification of patient notes with recurrent neural networks. Journal of the American Medical Informatics Association, 2017, 24(3): 596–606.