

核医学中的机器学习:第二部分——神经网络与临床

谢宗能^{1,2}(译) 颜建华²(审校)¹上海理工大学医疗器械与食品学院 200093;²上海健康医学院、上海市分子影像学重点实验室 201318本文首次发表在 *The Journal of Nuclear Medicine*, 2021, 62(1): 22-29

Machine learning in nuclear medicine: part 2—neural networks and clinical aspects

Katherine Zukotynski¹, Vincent Gaudet², Carlos F. Uribe³, Sulantha Mathotaarachchi⁴, Kenneth C. Smith⁵, Pedro Rosa-Neto⁴, François Bénard^{3,6}, Sandra E. Black⁷¹Departments of Medicine and Radiology, McMaster University, Hamilton, Ontario, Canada; ²Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada; ³PET Functional Imaging, BC Cancer, Vancouver, British Columbia, Canada; ⁴Translational Neuroimaging Lab, McGill University, Montreal, Quebec, Canada; ⁵Department of Eliversity of Toronto, Toronto, Ontario, Canada; ⁶Department of Radiology, University of British Columbia, Vancouver, British Columbia, Canada; ⁷Department of Medicine (Neurology), Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada
Corresponding author: Katherine Zukotynski, Email: katherine.zukotynski@utoronto.ca
原文 DOI:10.2967/jnumed.119.231837

【摘要】 第一部分概述了机器学习(ML)在核医学中的应用,第二部分重点介绍神经网络。该文先用一个例子说明神经网络是如何工作的,并讨论其潜在的应用。重点关注 ML 在图像重建、低剂量 PET、疾病检测以及诊断和预后预测的最新进展。目前在核医学中应用的 ML 算法呈现多样化,迫切需要对算法设计和结果指标解读进行规范,该文给出了一个规范清单以供业界参考和遵循。

【关键词】 机器学习;核医学;神经网络

DOI:10.3760/ema.j.cn321828-20210401-00091

机器学习(machine learning, ML)在核医学中应用系列的第一部分^[1]概述了 ML 算法及其基本构成。尽管随机森林^[2,4]和支持向量机^[5,6]等传统 ML 算法的应用在不断增加,但以人工神经网络(artificial neural networks, ANNs)为代表的复杂 ML 算法正越来越受到关注。此外,基于放射组学的 ANNs 在核医学中的应用也越来越普遍。放射组学特征通常指的是从医学图像中提取定量数据,如通过手工或计算机自动提取的反映肿瘤异质性的纹理特征^[7]。第二部分对当今使用的最强大的 ML 模型之一的 ANNs 进行了详细说明。

一、第一部分的简要回顾

ANNs 是最先进的 ML 算法(图 1),通常应用于分类(离散输出)或回归(模拟输出)。尽管 ANNs 已经存在了几十年,但直到最近才在医学成像中被广泛应用,这主要得益于技术进步以及可获得用于训练的大型数据集。输入到 ANN 的数据是分步骤处理的,其中每一步都由一层神经元组成。神经元是产生输入数据的加权总和,并叠加一个偏差来计算结果的非线性变换的计算单元。每一层的输出数据处理传递到下一层,直到最后一层产生输出结果为止。每一个 ANN 的框架设计描述了各个神经元和层之间的关系。ANNs 通常是受监督的,其使用标签数据来学习权重和偏差,其可以很简单(仅包括几层和单个输出),也可以很复杂。越复杂的 ANNs 通常具有更大的功能,但计算成本更高。复杂的 ANNs 主要用于深度学习。设计一个最优复杂度的 ANN 来解决特定的任务和获得足够多的高质量输入数据具有挑战性。当今,ANNs 是核医学中最常用的 ML 算法之一。表 1 总结了常用的 ML 术语。

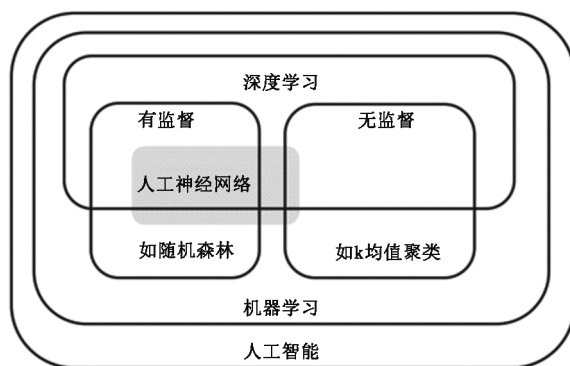


图 1 描述机器学习、人工智能和深度学习关系图。图例展示了简单的机器学习算法,如随机森林和 k 均值聚类。复杂的机器学习算法,如从神经网络延伸到有监督的深度学习,但不包含既不是有监督又不是无监督的算法(如强化学习算法)

二、ANNs:通过实例理解

笔者设计和训练了一个一层的神经网络来识别字母 X 的手写图像(图 2)。ANN 的输入是 8×8 的二维黑白图像或 64 像数阵列或向量($a_1 \sim a_{64}$),其中每个像素的值为 -1 (白色)和 1 (黑)(图 2A、2B)。输出(f)是一个介 0 和 1 之间的模拟数字,其代表输入为字母 X 的可能性,例如输入是 X 的图像,输出应该是 1,如果输入是另一个字母,输出应该是 0。这个 ANN 是采用一个数据集来训练,该数据集包括已经被标记为表示或不表示 X 的若干输入图像(图 2B)。通过将每个输入乘以相应的权重($w_1 \sim w_{64}$),乘积相加(假设所有偏差

均为 0),并通过称为激活函数(图 2C)的非线性函数(此处为 Sigmoid 函数)传递结果。

ANN 具有多种类型的激活函数,包括 Sigmoid 函数和修正线性单元(rectified linear unit, ReLU)(图 2D)。每个激活函数都会以某种方式约束输出,如 Sigmoid 函数将输出限制在 0 和 1 之间,而 ReLU 函数将得到一个非负数(图 2E)。这些非线性函数是优化 ANN 性能的关键。

在处理新图像之前,ANN 必须通过训练来学习权重值。为此,该算法使用一个成本(或损失)函数,该函数计算模型预测特定训练案例输出的接近程度。训练的最终目的是使成本函数最小化。常见的成本函数是计算训练输出(f_{train})和期望输出(f)之间的误差(E),这里是两者之间的绝对差(差值的平方和分类精度也是常见的成本函数):

$$E = |f - f_{train}| \dots\dots\dots \text{方程 1}$$

通过迭代训练来获得能够提供最佳性能($w_1 \sim w_{64}$)的值:重置初始化,将训练数据输入 ANN,计算误差函数,调整权重使 ANN 趋向更低成本,并给出新的训练参数。这个过程迭代进行,直到学习的权重得到满意的成本。

权重调整通常使用梯度下降算法,如随机梯度下降。计算成本函数的梯度,实质上是函数相对于每个权重的偏导数(即斜率)。一旦偏导数已知,权重就会朝着最陡的下降方向调整。然而,目前还没有关于需要调整相应权重的通用方法。调整太少,朝着最终目标进展不大;调整太多,可能导致输出下降。例如寻找函数 $E(x)$ 的一个极小点(图 3),如果从 A 点出发,应该向右移动;然而,一大步会走得太远(D 点)。

为了使 ANN 能够检测字母 X,需要推导出对每个权重的偏导数。首先,基于输入像素 a_i 和权重 w_i 来表示输出 f :

$$f = \frac{1}{1 + e^{-\sum_{j=1}^{64} a_j w_j}} \dots\dots\dots \text{方程 2}$$

对方程 1 求关于 w_i 得偏导数为:

$$\frac{\partial E}{\partial w_i} = \frac{(f_{train} - f)}{E} \frac{e^{-\sum_{j=1}^{64} a_j w_j}}{(1 + e^{-\sum_{j=1}^{64} a_j w_j})^2} a_i \dots\dots\dots \text{方程 3}$$

每个偏导数表示在每次迭代学习中相应的权重变化值。

现在用 24 个数据(手写样本;12 个字母 X 和 12 个其他

字母)训练 ANN。训练后,如图 4 所示权重为 8×8 的矩阵,神经网络中权重的灰度表示类似一个 X(图 4A)。直观的感觉是,因为权重反映的是概率图,所以权重的图像类似于被神经网络训练来检测的图像。

1.复杂的神经网络:层数和架构设计。具有深度学习能力的 ANNs 通常是多层结构^[8]。尽管对深度学习感兴趣的程序员可能会创建一个复杂的神经网络,但是每一层的任务通常不是由程序员预先定义的。相反,神经网络在任何情况下都像是一个黑匣子一样运作。

具有更多层的神经网络能够学习更复杂的事物,但需要更强大的计算能力,如需要图形处理单元(graphics processing units, GPUs)或云端远程服务器。有些神经网络超过 100 层和数百万权重则需要优化,其挑战在于建立一个 ANN,通过优化框架设计,用少量的操作来解决问题。

2.卷积神经网络(convolutional neural networks, CNNs)。CNNs 是 ANNs 的一种,其中层的结构采用了卷积核,这种结构适用于图像处理。卷积是一种常见的数学函数,核指的是一个权重矩阵,可以预先设置,也可以用更常见的方式通过 CNNs 训练来获取。CNNs 将一系列医学图像(通常是单模或多模)作为输入,然后执行操作,计算不同层的权重和偏差,并且基于期望的输出优化参数以使成本最小。通常,没必要输入感兴趣区或特征。虽然简单的 ML 算法可以在像素级处理图像,但 CNNs 在复杂决策方面具有更强大的能力,如在分类准确性方面。

图 5 介绍了 2 种常见的 CNNs 架构设计。对于输出为图像重建或分割等的应用来说,通常使用编码器-解码器体系结构的变体(图 5A)。编码器在逐步过程中减少输入数据以识别部件或特征。这可以通过使用表 1 或图 6 中定义的步距(stride)或池化(pooling)来实现。然后,解码器使用逐步(stepwise)过程从特征构建输出图像,其中可能包括内插或上采样以提高分辨率。遵循此样式的架构包括 U-Net(二维数据)和 V-Net(三维数据)。对于输出为疾病检测分类(如疾病存在或不存在)的应用,ANNs 可能仅具备编码器阶段,在此阶段中,输入数据在逐步过程中减少,从而导致输出分

表 1 神经网络术语

术语	说明	注释
全连接层(fully connected layer)	每个层的输入都被用于计算层的每个输出	图2C显示了具有64个输入和1个输出的全连接层;尽管输出数据点的数量少于输入数据点的数量,但这不是必须的
核(kernel)	卷积神经网络(CNN)中数字矩阵,通常是通过训练数据集进行训练获得的	3×3核或3×3×3核比较常见
步长(stride)	表示内核每次在CNN中处理图像时跳过的像素数的个数	图6表明了步长;输出图像比输入图像像素少,输出图像采用低维矩阵表示
池化(pooling)	CNN中通过对局部区域进行平均或取最大值来降低图像分辨率的操作	池化层,通过128×128矩阵图像作为输入,将输入矩阵分成2×2块,然后选出其中最大的数字,从而产生64×64矩阵图像作为输出
每秒浮点运算次数(FLOP)	FLOP代表浮点运算,代表计算能力的度量	与网络相关的FLOPs通常是指网络运行训练后所需的计算能力;在图2中,64次乘法和63次两两相加,代表127次FLOPs(省略Sigmoid函数);CNN可能需要数十亿次的FLOPs,而简单的机器学习算法,如随机森林或支持向量机可能也需要数千次

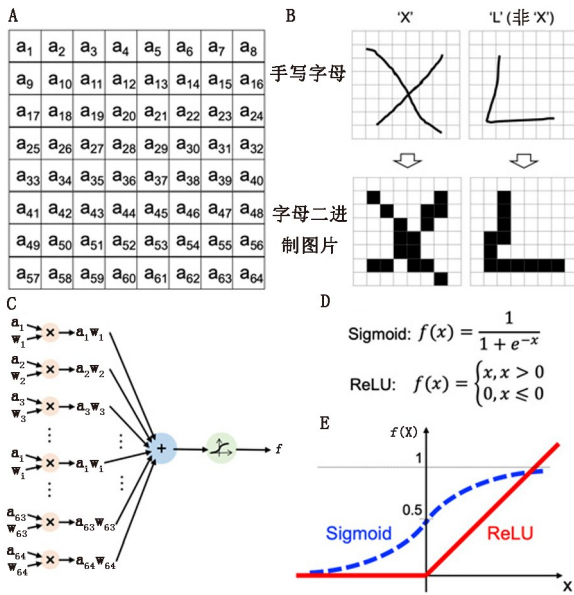


图 2 A. 输入: 8×8 像素矩阵 (a₁~a₆₄), 每个像素 -1 (白色) 或 +1 (黑色)。B. 映射到二进制的输入手写字母训练数据集。C. 单层神经网络, 其输入如 A 所示, 64 个权重 (w₁~w₆₄), Sigmoid 激活函数以及 1 个输出 (f), 输出是 0~1 的模拟数, 反映输入为 X 的概率。D. 激活函数的数学表达式。E. 显示 Sigmoid (蓝色) 和修正线性单元 (ReLU; 红色) 的图表; 对于 Sigmoid, 输出被限制在 0 和 1 之间; 对于 ReLU, 负输入被置为 0

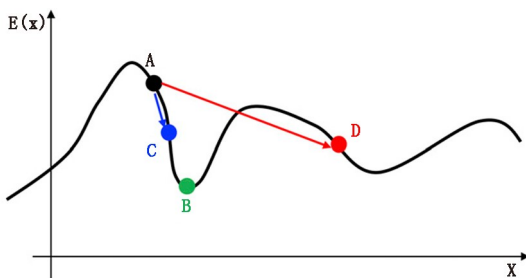


图 3 从点 A 开始希望找到函数 E(x) 中的最低点 (标记为 B)。假设知道 E(x) 在 A 点的斜率, 基于梯度的搜索建议向右移动较小的点, 理想的情况下, 希望函数搜索小步到 C, 而不是大步到 D

类 (图 5B)。Res-Net 就是这样一个框架。

3. 硬件方面。ANNs 通常使用 Matlab 或 Python 等软件进行编程, 但是运行速度主要取决于硬件。简单的 ANNs 可以在标准笔记本电脑上运行, 但更复杂的 ANNs 通常需要强大的硬件。GPU 是目前解决 ANNs 效率最有效的硬件解决方案, 因为它们能够同时执行许多简单的计算, 从而提高速度; 且 GPU 能执行卷积操作的所有计算。但当 ANNs 变得复杂时, 靠单个 GPU 不能解决问题, 这就需要计算集群 (通过由 GPU 组成的超级计算机)。

三、神经网络在核医学中的应用

ANNs 正被广泛应用于核医学领域。ANNs 可以帮助实现低剂量扫描获取标准剂量图像, 以及提高散射和衰减校正^[9-18]。ANNs 还可以帮助疾病检测和分割^[19-26]、疾病诊断和预后预测^[27-32]。

1. 神经网络在图像重建和低剂量 PET 显像中的应用。噪

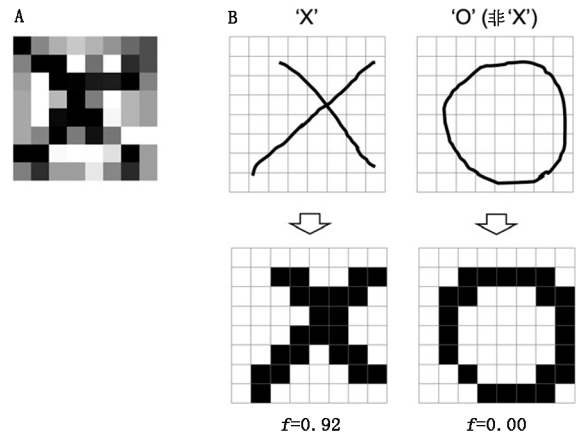


图 4 A. 经过 24 个训练案例后的权重矩阵 (核) 灰度图像, 颜色越深的像素权重越大, 注意到灰度图像类似于 X。B. 输入到训练过的神经网络中, 手写字母图像的新例子表明, 被识别为字母 X 的可能性为 92% (神经网络输出 f=0.92), O 被判定为不是 X 的可能性为 0% (f=0.00)

声是核医学显像过程中固有的信号, 低剂量显像技术或减少图像采集时间可能会加剧信号噪声。在图像重建过程中, CNNs 可以生成比传统技术更高质量的图像, 并改善噪声图像的感知质量。

Hägström 等^[11] 阐述了一种用于图像重建的神经网络方法。其使用类似于图 5A 的编码器-解码器架构构建了一个 CNN, 通过使用模拟、模体和增强技术得到数据进行 PET 图像重建。该 CNN 的输入是用 288×289×1 矩阵表示的 PET 正弦图数据, 输出为 128×128×1 矩阵表示的图像数据。该编码通过使用减小内核大小和步长为 2 的卷积核, 以及批量归一化和 ReLU 等激活函数, 通过顺序分层减少输入数据, 而解码器使用顺序层对数据进行上采样, 以应用卷积核, 增加矩阵大小, 并应用包括批量归一化和 ReLU 在内的激活函数来产生最终的 PET 图像。同时研究了几种变化设计, 包括不同层数和内核大小等。与有序子集最大期望值迭代法或滤波反投影法等技术相比, CNN 能够产生具有更高质量的 PET 图像。

CNN 包括几个具有平行路径的层 (也称为并行通道), 利用特定的核来剖析某些图像特征, 然后通过一系列层组合特征信息以产生无噪声的输出图像。有时输入一个有噪声的图像, 而 CNN 被设计为将这些输入数据减少到一定程度的低分辨率图像, 识别如边缘和纹理的抽象特征, 然后逐步重建一个与输入图像相同分辨率的无噪声输出图像。这些 CNNs 通常使用有噪声输入图像和无噪声输出图像进行监督训练, 确保没有重大信息丢失或虚假信息增加。

Chen 等^[12] 使用具有 U-Net 结构的 CNN (类似于图 5A 中所示), 从 1% 的原始列表 PET 数据获得低剂量图像后合成全剂量¹⁸F-氟比他班 (Florbetaben) PET/MR 图像。合成图像的质量由 2 名医师以 5 分制进行主观评价, 且使用 Bland-Altman 图来进行标准摄取值 (standardized uptake value, SUV) 比较。结果表明, 与低剂量图像相比, 合成图像的图像质量指标有所改善, 淀粉样蛋白状态的准确性高, 扫描器内再现类似于全剂量图像。文献^[33] 对使用 CNN 来提升低剂量 PET 显像进行了很好的总结。

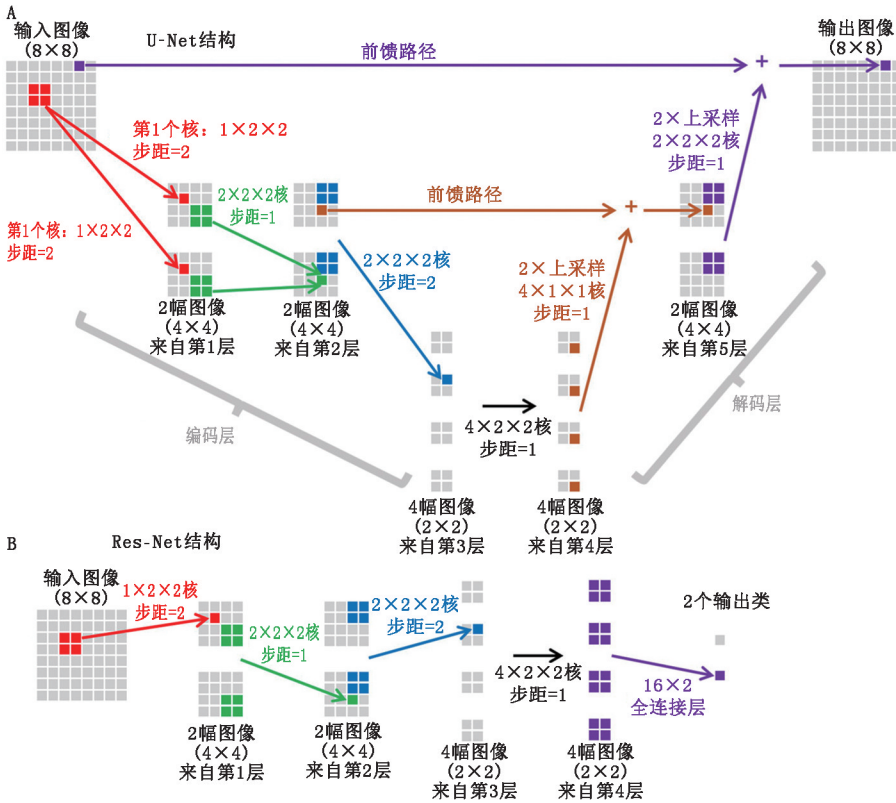


图 5 神经网络架构结构图解。A. 编码器-解码器设计有助于图像分割。在编码器中, 输入图像分辨率随着图像数量的增加而降低。第 1 层生成 2 幅图像, 首先使用核为 2×2, 进行步长为 2 的卷积运算, 然后采用不同大小的核。由于输入的是 1 幅图像, 将其表示为 1×2×2 核。第 2 层从 2 个输入图像中生成 2 个输出图像, 这些图像被视为体积 (volume), 同样使用 2 个不同的核, 记为 2×2×2, 步长为 1。第 3 层应用 4 个不同的 2×2×2 核, 步长为 2, 生成 4 张输入到第 4 层的图像。在解码器中, 上采样产生更高分辨率的图像, 因此卷积神经网络 (CNN) 的输入和输出分辨率是相似的。前馈路径添加来自较早层的数据, U 型结构命名为 U-Net。B. 编码器设计有助于疾病检测。在连续的图层上, 图像分辨率降低, 以识别编码搭配特征映射中的特征。最后一层通常是全连接层; 显示的 2 个输出分别使用来自上一层的每个像素的加权和, Res-Nets 就是这样的例子

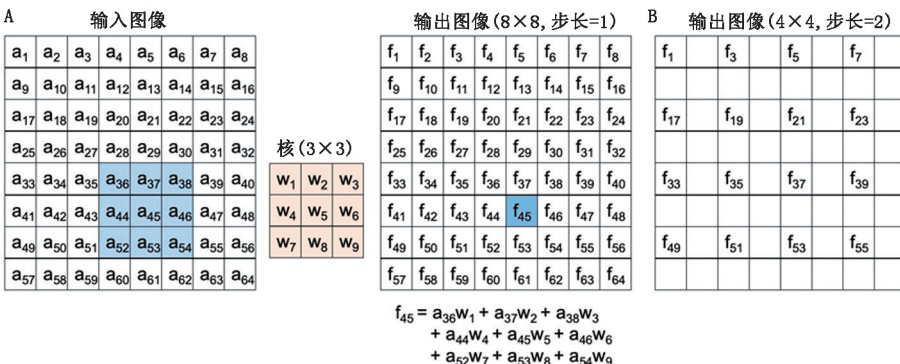


图 6 说明步长原理。A. 对输入 8×8 矩阵进行卷积层处理, 采用 3×3 核 (权重为 w₁~w₉)。输出 8×8 矩阵中的每个像素都是通过将近最接近相应输入像素的 9 个相邻像素乘以各自的内核权重来计算。示例为输出像素 f₄₅。B. 使用步长为 2, 在 2 个维度上计算隔一个像素值, 从而得到 4×4 输出图像

2. 神经网络在疾病检测和分割中的应用。神经网络的一个常见应用是疾病检测和分割, 如量化疾病程度。其本质上是一个像素分类问题; 每个像素必须被标记为正常或异常, 并加入到所属区域 (如肝脏或脾脏)。通常, 输出图像与输入图像具有相同的分辨率, 并叠加神经网络提取的特征信息从而得到分割图

像。与去噪类似, 输入和输出图像具有相同的分辨率, 分别使用原始图像和分割图像的组合来进行监督训练。几篇关于使用神经网络进行病变检测和分割的研究都使用了 U-Net, 但采用了不同的架构设计^[20-26]。

Zhao 等^[26] 采用 CNNs 对⁶⁸Ga-前列腺特异膜抗原 (prostate specific membrane antigen, PSMA)-11 PET/CT 图像疾病部位进行自动分割, 以判断某个体素是否为病灶。总体框架由 2 个串联的组件组成: 3 个并联 CNN 路径 (每个路径旨在检测 3 不同平面中的 1 个病变) 以及基于每个平面 CNN 的中间结果来进行结果投票。该 CNNs 具有一个 U-Net 结构, 该结构由一个编码堆栈和一个解码堆栈组成, 解码堆栈将特征映射和原始图像融合在一起, 其结构类似于图 5。编码堆栈包括 3×3 卷积, 2×2 最大池化, 步长为 2, 用于下采样、ReLU 和批量归一化; 解码堆栈采用核大小为 2×2、步长为 2 的转置卷积、级联运算、ReLU 和批量归一化的 3×3 卷积来合成信息。在 CNN 的最后一层, Sigmoid 函数将特征映射到分割概率图。戴斯 (Dice) 相似系数用于评估解剖分割的准确性。193 例耐药的转移前列腺患者的⁶⁸Ga-PSMA-11 PET/CT 扫描被随机分为 130 个训练集和 63 个测试集, 所有病灶都由人工勾画获取 (即 1 003 个骨病变和 626 个淋巴结病变等)。采用 5 倍交叉验证进行优化, 使用人工标注的图像作为标准, 当重叠率超过 10% 时, 就被认为病变。骨病灶的检测准确性、灵敏度和 F1 评分 (准确性和灵敏度的调和平均值) 分别为 0.99、0.99、0.99; 淋巴结上述指标分别为 0.94、0.90 和 0.92。图像分割精度低于病灶检测精度。整个模型对骨和淋巴结病变的平均 Dice 相似系数分别为 65% 和 54%, 阳性预测值分别为 80% 和 67%, 特异性分别为 61% 和 55%。

尽管 ANNs 有望用于病灶检测和图像分割且对临床实践有很大的影响, 但仍需通过专家来进行人工评估。

3. 神经网络在疾病诊断和预后预测中的应用。ANNs 可

以辅助疾病诊断和预后预测^[27-31]。通常只需要输入小部分图像或数据,而输入全分辨率图像或多个数据源的模型会逐渐减少数据的使用,以便通过最后一层提取诊断或结果。这些都是分类问题,其训练为有监督,且经常使用 Res-Net 架构设计,如图 5B 中所示。虽然早期的结果令人满意,但目前还缺乏强有力的证据。Nagendran 等^[34]的综述发现都存在类似问题,包括 1 项与乳房超声有关的随机临床试验和 2 项调查颅内出血的非随机前瞻性研究。这一领域的研究才开始,其对临床的影响还未知。

Mayerhoefer 等^[32]将神经网络用于预测,研究了¹⁸F-脱氧葡萄糖 PET/CT 的影像学参数是否可以单独或结合临床、实验室和生物学参数预测套细胞淋巴瘤患者的 2 年无进展生存 (progression-free survival, PFS)。该研究采用了多层前馈神经网络,反向传播学习算法^[8]和 logistic 回归分析进行特征选择。整个神经网络至少有 1 个隐藏层,且每个隐藏层至少有 3 个神经元,但没有具体阐述。输入包括对单个放射学特征权重的估计,并且分类操作重复进行 5 次。这个数据包括 107 个¹⁸F-FDG PET/CT 扫描,对接受治疗的套细胞淋巴瘤进行基线扫描和随访,直到病情进展、死亡或至少 2 年存活期。所有病例被随机分成 75 个训练病例和 32 个验证病例,每个分类步骤重复进行。病灶的轮廓勾画采用半自动方法,采用如下参数进行分析:最大 SUV,平均 SUV,SUV 峰值,病灶糖酵解总量以及由三维灰度共生矩阵中提取的 16 个纹理特征。结果检测包括受试者工作特征 (receiver operating characteristic, ROC) 曲线下面积和分类准确性。尽管放射学特征与绝对 PFS (以月计) 相关性没有统计学意义,但 2 年 PFS 与平均 SUV ($P=0.022$) 和熵 ($P=0.034$) 相关。当将平均 SUV 和熵值输入神经网络时,受试者 2 年 PFS 预测的 ROC 曲线下面积为 0.70~0.73 (中位数为 0.72),训练集的分类准确性为 71.0%~76.7% (中位数为 74.4%),验证集的分类准确性为 70.6%~86.8% (中位数为 74.3%),与其他临床、实验室或生物数据相结合后的分类准确性进一步提高。

4. 共同主题与号召。最后笔者给出一些关于神经网络在核医学中应用的结论。首先,良好的神经网络性能通常是在少于 10 层的情况下实现的。许多研究使用较少患者数据 (20~200),通过数据增强技术,得到更多的训练或交叉验证数据集,或者使用模拟软件生成数据。其次,ML 算法的计算成本很少被提及,但不应被忽视,因为其直接影响重复性和临床实用性。那些描述算法结构的论文经常省略关键信息,使得读者几乎不可能准确重现模型。浮点操作是计算机科学家和工程师常用的成本度量,但很少有这方面报道。常规临床实践中很少甚至没有对应用进行评估的随机临床试验。

现正处于将 ML 应用于核医学的早期阶段,需要来自不同业界的专家共同制定该领域发展标准模板和注意事项。这将使得评估算法的有效性、成本和适用性变得更加方便和易公开。如果掌握了详细信息,就可以用图表表示输入、算法复杂度和输出的度量,以建立对特定任务最有效的算法。图 7 给出了一个概念图,如果有了算法的标准化细节,则可以绘制该概念图。该图以低剂量 PET 为例,绘制百分比剂量与 ML 计算成本关系图。任何报告剂量、计算成本和算法种类的研究都可以作为图表上的一个点。随着更多数据累积,将看到趋势线的出现,如所示的虚线表示相同算法种类的恒定分类精度,这样可以推断算法能力的界限。Minarik 等^[10]发表关于 CNN 在各种图像噪声水平 (类似于百分比剂量) 下的性能结果,虽然 CNN 的表现很好,但没有给出相应的计算成本,而且很难准确复制。

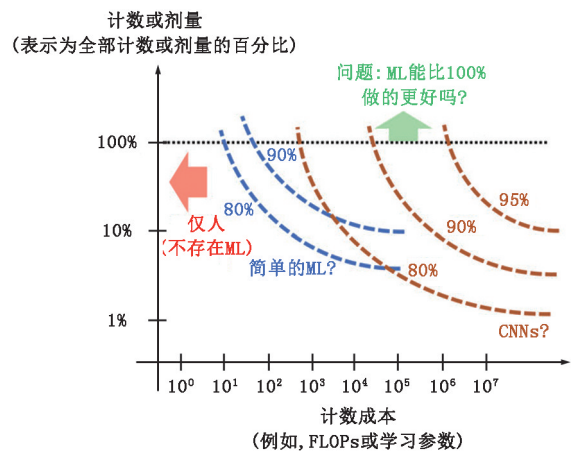


图 7 显示分类准确性 (虚线曲线) 和计数可能受机器学习 (ML) 算法计算成本 (以及学习复杂任务的能力) 的影响。这样的图表要求研究人员提供有关 ML 实现的具体细节。CNNs 为卷积神经网络, FLOP 为浮点运算

为了深入了解最适合给定任务的算法,以及实现所需输出的计算成本,建议使用列表来报告 ML 算法。表 2 提供了所期望的 5 个要点。

四、结论

神经网络在核医学中变得越来越普遍,且新的算法一直在涌现。此外,随着共享图像数据库的不断构建,将会有更多的数据用于训练、验证和测试中。但在实际使用的过程中仍然存在很多问题,特别是围绕数据收集、身份识别和所有权的道德和隐私问题。在某些情况下,将算法下载到多个站点可能比将多站点数据上传到公共数据库更容易。另外,

表 2 针对机器学习算法相关的建议清单

问题	指标度量	注释
机器学习算法?	机器学习算法种类	卷积神经网络 (CNN)、随机森林、支持向量机等
架构细节?	取决于算法	对于 CNN, 算法层数、核大小和步长, 并展示足够细节完整框图, 模型能完成独立重建
计算成本?	参数数量、浮点运算	建议咨询计算专家, 类似于临床实验的咨询统计学家, 但作者也可以自行计算
数据?	训练、验证、测试	数据类型, 验证/测试的数量, 交叉验证的使用, 数据源 (单个机构的数据训练的算法无法在另一个机构数据中表现良好)
性能指标?	分类准确性、剂量减少等	提供关键性能数字结果, 例如分类准确性; 最终, 对于给定的指标进行标准化

ML 算法指标的标准化对于临床应用大有帮助。表 2 给出了 5 个关键指标,并针对如何编译数据以生成图像以及针对不同的应用选用合适的机器算法给出了相应的建议。

参 考 文 献

- [1] Uribe CF, Mathotaarachchi S, Gaudet V, et al. Machine learning in nuclear medicine; part 1-introduction. *J Nucl Med.* 2019;60:451-458.
- [2] Zukotynski K, Gaudet V, Kuo PH, et al. The use of random forests to classify amyloid brain PET. *Clin Nucl Med.* 2019;44:784-788.
- [3] Nuvoli S, Spanu A, Fravolini ML, et al. ¹²³I-metaiodobenzylguanidine (MIBG) cardiac scintigraphy and automated classification techniques in Parkinsonian disorders. *Mol Imaging Biol.* 2020;22:703-710.
- [4] Perk T, Bradshaw T, Chen S, et al. Automated classification of benign and malignant lesions in ¹⁸F-NaF PET/CT images using machine learning. *Phys Med Biol.* 2018;63:225019.
- [5] Nicastro N, Wegrzyk J, Preti MG, et al. Classification of degenerative parkinsonism subtypes by support-vector-machine analysis and striatal ¹²³I-FP-CIT indices. *J Neurol.* 2019;266:1771-1781.
- [6] Kim JP, Kim J, Kim Y, et al. Staging and quantification of florbetaben PET images using machine learning: impact of predicted regional cortical tracer uptake and amyloid stage on clinical outcomes. *Eur J Nucl Med Mol Imaging.* 2020;47:1971-1983.
- [7] Mayerhoefer ME, Materka A, Langs G, et al. Introduction to radiomics. *J Nucl Med.* 2020;61:488-495.
- [8] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436-444.
- [9] Xiang L, Qiao Y, Nie D, An L, Wang Q, Shen D. Deep auto-context convolutional neural networks for standard dose PET image estimation from low-dose PET/MRI. *Neurocomputing.* 2017;267:406-416.
- [10] Minarik D, Enqvist O, Trägårdh E. Denoising of scintillation camera images using a deep convolutional neural network; a Monte Carlo simulation approach. *J Nucl Med.* 2020;61:298-303.
- [11] Häggström I, Schmidlein CR, Campanella G, Fuch TJ. DeepPET: a deep encoder-decoder network for directly solving the PET image reconstruction inverse problem. *Med Image Anal.* 2019;54:253-262.
- [12] Chen KT, Gong E, de Carvalho Macruz FB, et al. Ultra-low-dose ¹⁸F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. *Radiology.* 2019;290:649-656.
- [13] Gao F, Shah V, Sibille L, Zuehlsdorff S. An AI system to determine reconstruction parameters and improve PET image quality [abstract]. *J Nucl Med.* 2018;59(suppl 1):31.
- [14] Hwang D, Kim KY, Kang SK, et al. Improving the accuracy of simultaneously reconstructed activity and attenuation maps using deep learning. *J Nucl Med.* 2018;59:1624-1629.
- [15] Leynes AP, Yang J, Wiesinger F, et al. Zero-echo-time and Dixon deep pseudo-CT (ZeDD CT): direct generation of pseudo-CT images for pelvic PET/MRI attenuation correction using deep convolutional neural networks with multipara-metric MRI. *J Nucl Med.* 2018;59:852-858.
- [16] Spuhler KD, Gardus J III, Gao Y, DeLorenzo C, Parsey R, Huang C. Synthesis of patient-specific transmission data for PET attenuation correction for PET/MR neuroimaging using a convolutional neural network. *J Nucl Med.* 2019;60:555-560.
- [17] Torrado-Carvajal A, Vera-Olmos J, Izquierdo-Garcia D, et al. Dixon-VIBE deep learning (DIVIDE) pseudo-CT synthesis for pelvic PET/MR attenuation correction. *J Nucl Med.* 2019;60:429-435.
- [18] Gong K, Guan J, Kim K, et al. Iterative PET image reconstruction using convolutional neural network representation. *IEEE Trans Med Imaging.* 2019;38:675-685.
- [19] Lindgren Belal S, Sadik M, Kaboteh R, et al. Deep learning for segmentation of 49 selected bones in CT scans; first step in automated PET/CT-based 3D quantification of skeletal masses. *Eur J Radiol.* 2019;113:89-95.
- [20] Gsaxner C, Roth PM, Wallner J, Egger J. Exploit fully automatic low-level segmented PET data for training high-level deep learning algorithms for the corresponding CT data. *PLoS One.* 2019;14:e0212550.
- [21] Huang B, Chen Z, Wu PM, et al. Fully automated delineation of gross tumor volume for head and neck cancer on PET-CT using deep learning: a dual-center study. *Contrast Media Mol Imaging.* 2018;2018:8923028.
- [22] Zhao X, Li L, Lu W, Tan S. Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network. *Phys Med Biol.* 2018;64:015011.
- [23] Hatt M, Laurent B, Ouahabi A, et al. The first MICCAI challenge on PET tumor segmentation. *Med Image Anal.* 2018;44:177-195.
- [24] Bi L, Kim J, Kumar A, Wen L, Feng D, Fulham M. Automatic detection and classification of regions of FDG uptake in whole-body PET-CT lymphoma studies. *Comput Med Imaging Graph.* 2017;60:3-10.
- [25] Xu L, Tetteh G, Lipkova J, et al. Automated whole-body bone lesion detection for multiple myeloma on ⁶⁸Ga-pentixafor PET/CT imaging using deep learning methods. *Contrast Media Mol Imaging.* 2018;2018:2391925.
- [26] Zhao Y, Gafita A, Vollnberg B, et al. Deep neural network for automatic characterization of lesions on ⁶⁸Ga-PSMA-11 PET/CT. *Eur J Nucl Med Mol Imaging.* 2020;47:603-613.
- [27] Commandeur F, Goeller M, Razpour A, et al. Fully automated CT quantification of epicardial adipose tissue by deep learning: a multi-center study. *Radiol Artif Intell.* 2019;1:e190045.
- [28] Eisenberg E, Commandeur F, Chen X, et al. Deep learning-based quantification of epicardial adipose tissue volume and attenuation predicts major adverse cardiovascular events in asymptomatic subjects. *Circ Cardiovasc Imaging.* 2020;13:e009829.
- [29] Hartenstein A, Lubbe F, Baur ADJ, et al. Prostate cancer nodal staging; using deep learning to predict ⁶⁸Ga-PSMA-positivity from CT imaging alone. *Sci Rep.* 2020;10:3398.
- [30] van Velzen SGM, Lessmann N, Velthuis BK, et al. Deep learning for automatic calcium scoring in CT: validation using multiple cardiac CT and chest CT protocols. *Radiology.* 2020;295:66-79.
- [31] Huang Y, Xu J, Zhou Y, et al. Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network. *Front Neurosci.* 2019;13:509-520.
- [32] Mayerhoefer ME, Riedl CC, Kumar A, et al. Radiomic features of glucose metabolism enable prediction of outcome in mantle cell lymphoma. *Eur J Nucl Med Mol Imaging.* 2019;46:2760-2769.
- [33] Zaharchuk G. Next generation research applications for hybrid PET/MR and PET/CT imaging using deep learning. *Eur J Nucl Med Mol Imaging.* 2019;46:2700-2707.
- [34] Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: a systematic review of design, reporting standards, and claims of deep learning studies. *BMJ.* 2020;368:m689.