

脑血管病中的机器学习

Yannan Yu and David Yen-Ting Chen

Re-edited by Huandi Shi

1 引言	2
2 缺血性中风	3
2.1 大血管闭塞 (LVO) 的诊断	4
2.2 预测中风发作时间	8
2.3 中风病灶分割	8
2.4 预测未来的中风病灶	9
2.5 预测出血性转化	10
2.6 预测中风临床结局	11
2.7 预测脑血流量 (CBF) 和脑血管储备 (CVR)	11
3 出血性中风 (卒中) 或颅内出血	12
4 脑血管畸形	13
5 颅内动脉瘤	13
5.1 动脉瘤检测中的困难	14
5.1.1 颅内动脉瘤检测的人工智能算法	14
5.2 动脉瘤风险评估的困难	15
5.2.1 基于人工智能的动脉瘤风险预测模型	15
6 脑小血管病	16
6.1 脑小血管病的影像学特征	16

6.2 白质高信号分割	17
6.2.1 基于深度学习的白质高信号分割方法	18
6.3 脑微出血 (CMB) 检测	20
6.4 腔隙性病灶检测	20
6.5 血管周围空间定量	21
6.6 总小血管病负荷	21
7 结论	22

摘要

脑血管疾病指的是一组影响大脑内血液流动和血管的病症。它是全球范围内导致死亡和致残的主要原因之一，对社会造成了显著的社会经济负担。近年来，脑血管疾病的研究进展迅速，使得患者的诊断和管理得到了改善。机器学习在进一步提高这些疾病的临床护理方面具有很大的潜力。在本章中，我们将简要介绍脑血管疾病的基本信息，并总结机器学习在改善研究和患者护理方面最具潜力的一些领域。

具体而言，我们将涵盖以下脑血管疾病：中风（包括缺血性和出血性）、脑微出血、脑血管畸形、颅内动脉瘤和脑小血管病（白质高信号、腔隙性梗死、血管周围间隙）。

1 引言

脑血管疾病是一组影响大脑血管和脑部血液循环的疾病。中风是脑血管疾病最常见的表现。大多数中风是缺血性的，由于血流减少导致脑组织损伤和神经功能障碍。较少见的是出血性中风，造成血液从脑血管渗出到脑组织（颅内出血）或脑组织周围的空间（蛛网膜下腔和硬膜下出血）。出血性中风可能导致灾难性损伤，原因是颅内压升高、脑组织灌注减少和正常脑组织受损。2019 年，全球因脑血管疾病导致的死亡人数达到 660 万人；其中 300 万人死于缺血性中风，290 万人死于脑内出血，40 万人死于蛛网膜下腔出血 [1]。中风是全球第二大死因，占所有死亡人数的 11.6%，并且是导致死亡和残疾的第三大原因，造成了 143 百万个伤残调整生命年 [2]。脑小血管病包括影响大脑小穿通动脉、毛细血管和小静脉的一系列疾病。它有广泛的临床表现，约造成 25% 的中风，并导致约 45% 的痴呆病例 [3]。脑小血管病在老年人群中高度流行，50 岁人群中约 5% 受其影响，而 90

岁以上人群中几乎 100% 受其影响 [3]。颅内动脉瘤 (IA) 是由于大脑血管的膨出造成的；如果动脉瘤破裂，可能导致灾难性的蛛网膜下腔出血，死亡率为 23-51% [4, 5]，并有 30-40% 的患者会有永久性残疾 [4, 6]。动静脉畸形 (AVM) 是由于大脑中血管缠结而导致的，绕过正常脑组织；动静脉畸形可能引起出血和癫痫。

脑血管疾病通常通过影像学检查进行诊断，某些脑血管疾病的治疗基于影像引导。常见的影像学检查方法包括计算机断层扫描 (CT)、磁共振成像 (MRI) 和数字减影血管造影 (DSA)。CT 能够快速检查脑组织和脑血管；本章将提到的一些 CT 协议包括无对比 CT、CT 血管造影 (CTA) 和 CT 灌注 (CTP)。无对比 CT 是诊断颅内出血的首选检查，也是缺血性中风初步分流的首选检查。然而，缺血性中风在无对比 CT 上的表现主要取决于中风的年龄，变化范围从 0-6 小时内无变化或细微变化到 24 小时后明显低密度。增强 CT 根据详细协议可以突出血管结构，称为 CTA，常用于诊断缺血性中风、动脉瘤或动静脉畸形中的动脉闭塞。增强 CT 还可以计算脑部血流灌注状态，称为 CTP，常用于缺血性中风的分流。MRI 有多种序列，可以为医学诊断提供特定的组织外观。本章将提到的一些序列包括扩散加权成像 (DWI)，其测量水分子运动的限制，对中风中的受损组织非常敏感。灌注加权成像 (PWI) 和动脉自旋标记 (ASL) 均测量脑灌注，但 PWI 需要注射对比剂，而 ASL 则不需要。两者常用于缺血性中风。梯度回波 (GRE) 和敏感加权成像 (SWI) 对铁和钙沉积非常敏感，常用于血液产品检测，且可用于检测出血和小血管病。T2 加权流体衰减反转恢复 (FLAIR) 常用于检测大于 6 小时的中风病灶和小血管病。对于 CT 灌注和 MR 灌注，可以计算定量灌注图，以估计血流灌注状态，常见的包括脑血流量 (CBF)、脑血容量 (CBV)、残余功能最大值时间 (Tmax) 和平均通行时间 (MTT)。数字减影血管造影 (DSA) 是一种荧光透视技术（类似于 X 射线），用于可视化血管，常用于诊断和治疗动脉瘤、缺血性中风的动脉闭塞以及一些动静脉畸形。

机器学习在优化脑血管疾病护理方面具有很大的潜力，能够利用临床和影像数据改善或加速诊断并提供预后评估。

2 缺血性中风

大约 87% 的中风为缺血性中风，13% 为出血性中风 [1]。缺血性中风是由于大脑某部分的血液供应减少或缺失，通常由脑动脉的阻塞或狭窄引起，导致局部脑组织损伤和神经功能丧失。缺血性脑损伤与时间密切相关 [7]。目前唯一推荐的治疗方法是在症状出现后 4.5 小时内进行静脉溶栓治疗，以及在症状出现后 24 小时内进行血管内取栓治疗；这些治疗仅适用于特定类型的中风患者 [8]。急性中风治疗旨在重新通畅被阻塞的脑血管，恢复缺血或灌注不足的脑组织的血流，具体通过静脉药物来溶解血栓（溶栓治疗）或在罪犯动脉内机械性移除血栓（血管内取栓）。由于临床协议时间敏感且标准化，因此及时诊断缺血性中风和快速启动治疗是临床实践中的关键步骤 [7]。因此，基于机器学习的算法在急性缺血性中风护理中具有巨大潜力。图 1 展示了中风在临床环境中诊断和治疗的

基本流程示例。

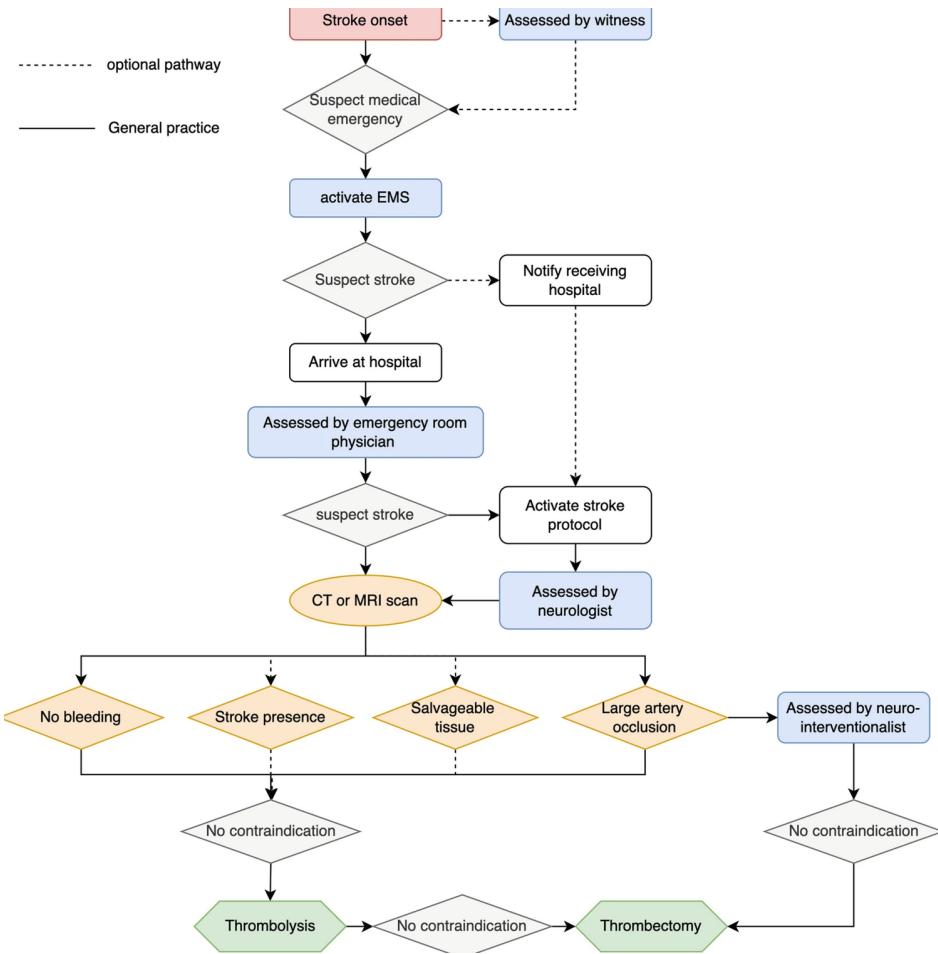


图 1: 卒中诊断和治疗的一般路径。实线表示常规做法；虚线表示可选路径。EMS 表示紧急医疗服务，CT 表示计算机断层扫描，MRI 表示磁共振成像。

本节将回顾研究机器学习在大血管闭塞 (LVO) 诊断、中风发作时间评估、中风病灶分割、中风结果和并发症预测中的应用。急性中风的常见影像学检查方法为计算机断层扫描 (CT) 和磁共振成像 (MRI)，用于中风的诊断和分流，数字减影血管造影 (DSA) 则用于中风的诊断和治疗。这些影像学检查的示例在图 2 和图 3 中展示。

2.1 大血管闭塞 (LVO) 的诊断

大血管闭塞 (LVO) 定义为近端颅内动脉的阻塞，占急性缺血性中风的约 24-46% [9]。诊断 LVO 是中风诊断和治疗考虑的重要步骤；患有 LVO 的患者是血管内取栓术的潜在候选者，这是目前最有效的疏通阻塞动脉的治疗方法 [8, 10, 11]。血管内取栓术是一种高度专业化的手术，所需的人员和设备并不广泛可用。患者通常需要从最初评估的医院转移到综合中风中心。在初步分诊期间，LVO 的自动检测可以加速急性中风协议和患者转移 [12]。

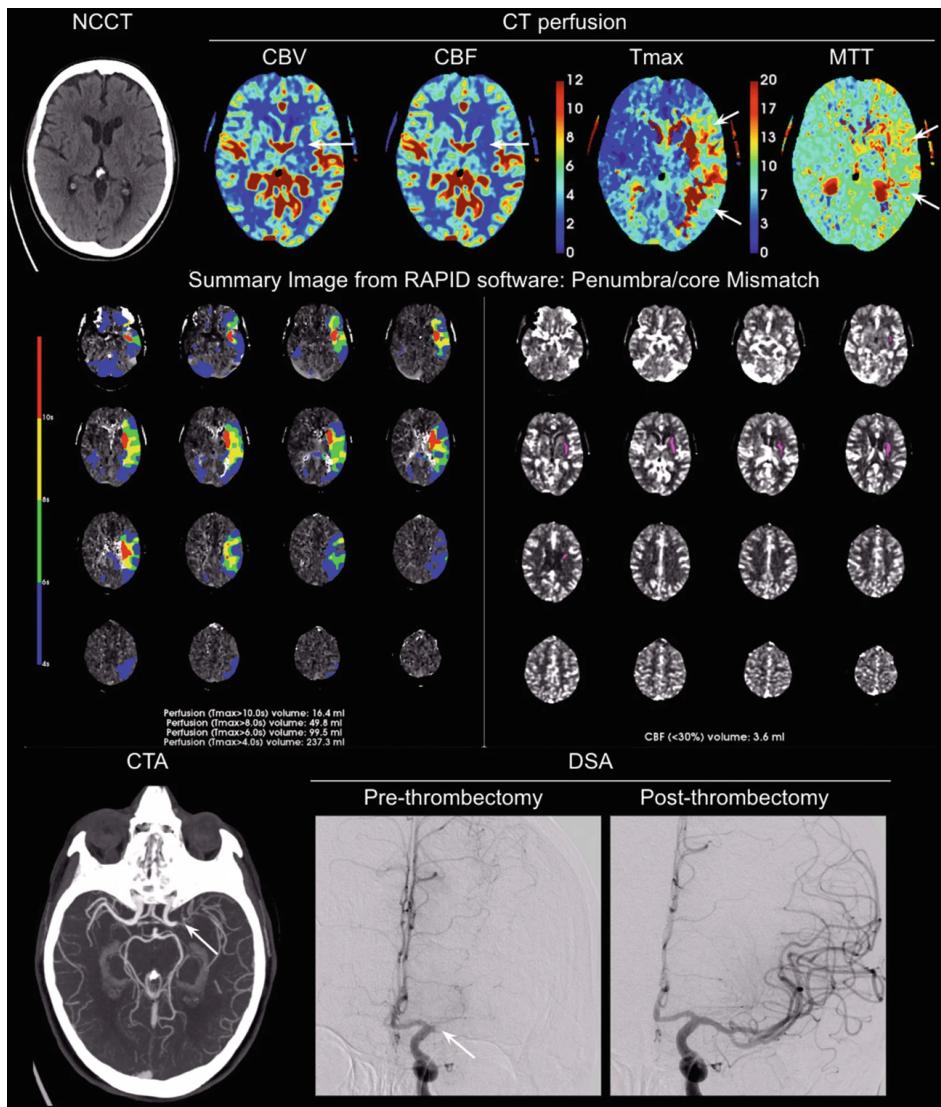


图 2: 常见的急性卒中 CT 扫描。本例显示左侧卒中（图像右侧），伴有大脑中动脉（M1 段）闭塞。平扫 CT (NCCT) 仅显示非常微妙的变化，而 CT 灌注显示大面积灌注缺损 (CBF 出现不对称低值，Tmax 和 MTT 出现不对称高值) 和小范围不可逆组织损伤。半暗带/核心不匹配是指延长的 Tmax 区域与降低的 CBF 区域的体积比；RAPID 软件的总结图像显示不匹配比率为 99.5 mL/3.6 mL。CTA 显示大脑中动脉主干 (M1 段) 闭塞。DSA 图像显示取栓术后动脉闭塞再通。NCCT 指非增强计算机断层扫描，CBV 指脑血容量，CBF 指脑血流量，Tmax 指组织残余功能的最大时间，MTT 指平均通过时间，CTA 指计算机断层扫描血管造影，DSA 指数字减影血管造影。

CT 血管造影 (CTA) 是快速、非侵入性诊断大血管闭塞的首选影像方式。多项研究利用机器学习展示了在 CTA 上识别 LVO 的可行性。Viz.ai 开发了一种商业化的 LVO 检测方法，通过 3D U-Net 的 CTA 血管分割和通过端点长度及 Hounsfield 单位 (CT 图像像素的标准单位) 值在大脑中动脉分支分割中的对比进行的大血管分类来实现。Yahav-Dovrat 等 [13] 在一项包含 404 例中风协议 CTA 的前瞻性队列中报告了该系统的性能。404 例中 72 例有 LVO，软件显示的敏感性为 82%，阳性预测值为 64%，阴性预测值为 96%。相对较低的敏感性和阳性预测值可能限制了该模型的临床

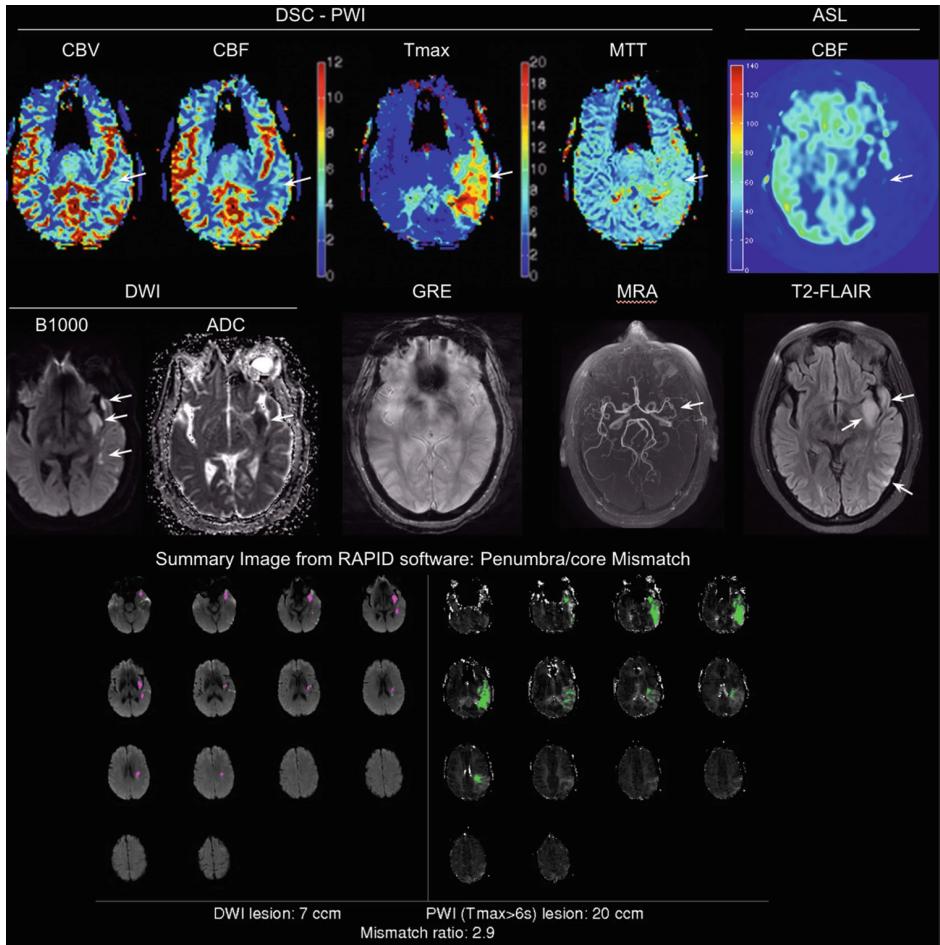


图 3: 急性卒中中常用的 MRI 序列。本例显示左侧卒中（图像右侧），伴有大脑中动脉分支（M2 段，下部分支）闭塞。DSC-PWI 和 ASL 显示灌注缺损（CBF 出现不对称低值，Tmax 和 MTT 出现不对称高值），其程度远大于 DWI 上观察到的不可逆组织损伤，且不匹配比率为 2.9。GRE 未显示出 blooming 效应（急性动脉内血栓的常见发现），MRA 显示左侧大血管闭塞（M2 段，白色箭头），而在卒中后 24 小时拍摄的 T2-FLAIR 显示卒中后受损的脑组织（白色箭头）。DSC-PWI 指动态敏感对比灌注加权成像，ASL 指动脉自旋标记，CBV 指脑血容量，CBF 指脑血流量，Tmax 指组织残余功能的最大时间，MTT 指平均通过时间，DWI 指扩散加权成像，ADC 指表观扩散系数，GRE 指梯度回波序列，MRA 指磁共振血管成像，T2-FLAIR 指 T2 加权液体衰减反转恢复。

实用性，因为急性缺血性中风筛查过程需要高敏感性。Stib 等 [14] 使用 270 例 LVO 患者和 270 例无 LVO 患者的多阶段 CTA 最大强度投影 (MIP) 图像训练卷积神经网络 (CNNs)。然后在一个包含 62 名患者的平衡数据集中测试模型，使用多阶段 CTA 的所有阶段显示出 100% 的敏感性和 77% 的特异性，优于单阶段 CTA 的 77% 敏感性和 71% 特异性。值得注意的是，RAPID 的一种非深度学习商业方法在独立验证队列中显示出优异的敏感性和特异性（均在 95% 以上）[15]。这些自动化技术已整合到许多中风护理系统的临床实践中，并且进一步针对特定中心人群的算法优化可能提高临床性能。

LVO 也可以从非血管造影图像中检测到，特别是非对比 CT，这比 CTA 更为广泛可用。CTA 需要

静脉注射造影剂，通常不适用于肾功能衰竭和/或对碘造影剂过敏的患者。You 等 [16, 17] 报告了一种使用 200 例临床数据和从 U-Net 瓶颈提取的非对比 CT 图像特征训练的 XGBoost 模型，该模型在 100 个测试案例中显示出 95.3% 的敏感性和 68.4% 的特异性。Olive-Gadea 等 [18] 报告了一种基于 DenseNet 和决策树的预测模型，以诊断非对比 CT 图像中的 LVO，显示出 83.1% 的敏感性和 85.1% 的特异性，优于基于国立卫生研究院（NIH）中风量表模型的性能。

数字减影血管造影（DSA）是一种用于大血管闭塞（LVO）的侵入性诊断方法，用于指导介入神经放射科医生进行血管闭塞的治疗。取栓治疗是在 DSA 的指导下进行的，以移除导致闭塞的血栓。然而，解读 DSA 图像需要介入神经放射学的高度专业化培训，并且在取栓过程中通常需要对治疗效果进行实时评估。脑梗死溶栓（TICI）评分是在取栓手术后对中风治疗效果进行 DSA 评估的标准。之前的研究报告称，TICI 的评估者间一致性较低 [19, 20]。在 DSA 研究中应用机器学习具有挑战性，因为 DSA 包含来自三维血管的二维投影图像，这些图像对 X 射线探测器平面的位置非常敏感，并且还包含使数据更类似于视频的时间信息。在解读 DSA 图像时，放射科医生关注的是与正常图谱相比的解剖差异、造影剂填充动脉的速度、造影剂填充毛细血管系统的范围以及造影剂从静脉的排出情况。

Ueda 等 [21] 收集了带有和不带有错配伪影的 DSA 图像，并应用 U-Net 和卷积补丁生成对抗网络架构作为生成器和判别器网络，以从错配的 DSA 中预测非错配的 DSA。Zhang 等 [22] 提出了一种用于跟踪和分割 DSA 中脑血管的 U-Net，这可能是构建诊断工具的第一步。由于 DSA 是带有时间信息的二维图像，研究采用了不同的策略将这些特征融入神经网络。Bhurwani 等 [23] 提出了一种用于取栓术后 DSA 图像的集成卷积神经网络，以预测再灌注状态。他们在诊断取栓术后再灌注方面实现了 90% 的敏感性和 74% 的特异性。Su 等 [24] 提出了一种包括阶段分类、运动校正和灌注分割的精心设计的算法，使用 ResNet-18 实现最终的 TICI 评分。他们在算法和人类读者之间实现了 90% 的一致性。值得注意的是，人类与人类之间的一致性为 89%。同一研究组的研究人员 [25] 还设计了一种复杂的网络，用于空间和时间特征提取，并预测取栓术中的并发症——血管穿孔。该模型以 0.83 的精度和 0.70 的召回率预测穿孔，其性能与人类专家读者相似。

除了通过影像学对大血管闭塞（LVO）进行分类外，研究还表明，基于临床评估预测 LVO 也是可行的，这可以为紧急医疗服务（EMS）直接运输到综合中风中心做好准备 [26-30]。Chen 等 [27] 使用 600 名患者（LVO 与非 LVO 比率为 1:1）的 NIHSS 分解评分、人口统计学数据、病史和风险因素作为输入，采用十倍交叉验证训练了人工神经网络（ANN）模型。该 ANN 模型的敏感性达到 0.807，特异性为 0.833。来自同一团队的 Wang 等 [26] 随后在 15,365 名患者中训练了 8 种机器学习模型，并在 4215 名患者上进行测试，使用他们的 NIHSS、人口统计学数据、病史和风险因素作为输入。他们的研究显示，随机森林模型表现最佳，达到 0.831 的 AUC，敏感性为 0.721，特异性为 0.827。

2.2 预测中风发作时间

在 14-27% 的中风病例中，症状起始时间是未知的 [31-33]。对于这些患者，识别可能的发作时间对于适当的治疗至关重要。确切地知道是否仍在静脉溶栓（4.5 小时内）或血管内治疗（如果存在大血管闭塞且非对比 CT 无广泛病灶则为 6 小时内，或如果存在大血管闭塞且灌注成像呈现目标不匹配则为 24 小时内）的治疗窗口内是关键。MRI 在估计中风持续时间方面起着重要作用。研究表明，液体衰减反转恢复 (FLAIR) 通常在中风发作后 3-6 小时发现缺血性病灶 [34, 35]，而弥散加权成像 (DWI) 则在中风后几分钟内即可检测到缺血性病灶。因此，FLAIR 和 DWI 之间的“错配”可用于确定中风起始时间 [36]。Lee 等 [37] 从 DWI 和 FLAIR 影像中捕获了 89 个向量特征，并训练了包括逻辑回归、支持向量机和随机森林在内的机器学习模型，以判断中风发作是否在 4.5 小时内。相比于人类读者，他们发现机器学习模型更具敏感性 (75.8% 对 48.5%， $p = 0.01$)，但特异性较低 (82.6% 对 91.3%， $p = 0.15$)。其他研究组也取得了类似的结果 [38]。过去并未研究灌注 MRI 用于确定中风起始时间。Ho 等 [39, 40] 使用自动编码器从灌注 MRI 中提取深度特征，以分类中风起始时间是否在 4.5 小时内（目前静脉注射组织型纤溶酶原激活剂 [tPA] 的时间窗口）。使用 DWI、表观扩散系数 (ADC)、FLAIR 和灌注加权图像作为输入，他们达到了 0.765 的 ROC AUC。这种方法在同一数据集中优于基于 DWI-FLAIR 的机器学习方法 (AUC 为 0.669) 和临床方法 (AUC 为 0.58)。使用影像学来确定中风发作时间可能会增加有资格接受时间限制中风治疗（如静脉溶栓）的患者数量 [31]。

2.3 中风病灶分割

非增强 CT 扫描是中风患者最常见的初始影像学检查。因此，相较于 MRI 数据集，CT 数据集通常更为常见且规模更大。然而，在 CT 上诊断早期中风或预测最终中风病灶通常比在 MRI 上更具挑战性，因为 CT 上与缺血性中风的早期急性期 (<6 小时) 相关的变化非常微妙，包括灰白质分界不清、深部核团低密度、皮质低密度以及相关实质肿胀和脑回消失。阿尔伯塔中风计划早期 CT 评分 (ASPECTS) 是一种基于非增强 CT 图像的早期急性期变化评估中风病灶存在的评分系统，评分范围为 0 到 10，其中 0 表示广泛的缺血性损伤，10 表示无缺血证据 [41]。当前指南建议对 ASPECTS 评分高的患者进行再灌注治疗 [8]，这意味着组织损伤较少，但正在进行的研究和试验正在调查对低 ASPECTS 中风患者进行治疗的益处 [42]。DWI/ADC 是识别早期中风病灶最常用且最准确的 MRI 序列（使用 ADC $620 \times 10^{-6} \text{ mm}^2/\text{s}$ 的阈值）。此外，MRI/CT 上的自动分割将有利于急性治疗决策，并使研究人员能够在更大规模上进行临床研究。

许多研究表明，机器学习可用于急性至亚急性期 CT 和 MRI 的中风病灶分割 [43-57]。Kuang 等 [58] 使用来自 157 名中风患者的非增强 CT 图像训练了一个随机森林分类器，以预测 CT 图像后 1 小时内扫描的 MRI 上的 ASPECTS 评分，并在 100 名患者上进行了测试。他们在 100×10 个 ASPECTS 区域中获得了 66.2% 的敏感性和 91.8% 的特异性，并在分类 ASPECTS >4 和 4 时达

到了 97.8% 的敏感性和 80% 的特异性。同一团队的 Qiu 等 [57] 使用相同的数据集，以 MRI 为基准，在非增强 CT 图像上分割早期中风病灶。他们提出了一种随机森林算法，结合了复杂的特征工程，包括距离特征、编码病灶位置特征的图谱，以及从独立数据集生成的 U-Net 病灶概率图作为输入。他们展示了预测的中风病灶体积与基准之间的良好相关性 ($r = 0.76$)，平均体积差异为 11 mL。两种用于自动 ASPECTS 评分的商业软件 (e-ASPECTS, Brainomix 和 Rapid ASPECTS, iSchemaView) 已经上市，并据报道其准确性不逊于甚至高于临床医生 [59-64]。

在 2015 年的缺血性中风病灶分割挑战赛 (ISLES) 中，提供了用于亚急性中风病灶分割的 MRI 序列训练和测试数据，包括 DWI 和 FLAIR。在此挑战中，3D CNN 达到了最高的病灶分割性能，Dice 相似系数 (DSC) 为 0.57 [45]。Chen 等 [43] 开发了一种两步法，从 DWI 中分割中风病灶，达到了 0.67 的 DSC。第一步使用编码器-解码器 CNN 进行病灶分割提议，第二步的 CNN 接收原始 DWI 图像和前一步输出的多尺度图像块作为输入，分类提议的分割为真或假。其他研究使用 2D 和 3D 编码器-解码器 CNN 也取得了类似结果 (DSC 0.64-0.76) [47-50]。ISLES 2018 挑战赛提供了急性中风 CT 灌注成像的训练和测试数据，以预测在 DWI 上定义的不可逆受损组织 [65]。排名最高的团队使用了具有空洞卷积算法的 3D 多尺度 U-Net，平均 DSC 为 0.51，平均绝对体积差异为 10.2 mL [66]。其他研究也取得了类似结果，但准确性不及表现最佳的团队 (DSC 0.44-0.49) [67, 68]。

上述方法需要对大量图像进行手动标注以用于训练，这既昂贵又限制了医学影像深度学习研究的规模。因此，Zhao 等 [52] 探索了在弱标注的中风分割数据集上使用急性 DWI 和 ADC 的半监督算法（结合 K-means 聚类和 CNN），达到了平均 DSC 为 0.64。Federau 等 [53] 使用带有合成中风病灶的 DWI 数据集进行 3D U-Net 分割，实现了 DSC 0.72。最近，Zhang 等 [51] 利用特征金字塔网络 [69] 和多平面（轴向、矢状和冠状平面）DWI 的 U-Net 进行病灶分割，获得了 DSC 0.62。由于放射科医生通常通过查看不同序列来解读 MRI，输入不同成像序列并“融合”信息的神经网络是提高诊断的重要研究方向。

Winzeck 等 [55] 提出训练 CNN 集成，而不是单个 CNN。作者采用了 ISLES 2015 挑战中性能最高模型的 CNN 结构，发现由五个 3D CNN 组成的集成在从 ADC、DWI 和 B0 图像中分割 DWI 病灶时比单个 CNN 更准确（中位 DSC 0.82 对比 0.79）。同一团队的 Wu 等 [44] 使用多中心、多厂商数据集（包含 ADC、DWI 和 B0 数据）训练了 CNN 集成，并发现其性能优于使用单中心数据集训练的模型，中位 DSC 为 0.86 (IQR 0.79-0.89)。虽然不同论文使用的测试数据集不同，无法直接比较模型性能，但这一章报告了迄今为止中风病灶分割的最高 DSC。

2.4 预测未来的中风病灶

与在单一时间点成像的中风病灶分割相比，使用基线 CT/MRI 在随访图像上分割最终病灶或出血性转化，是一种预测患者未来临床和影像学结果的方法。特别是，能够预测个体对治疗反应的方法（例如，在有治疗和无治疗情况下预测未来结果）可以帮助确定治疗是否对该个体有益。

2016 和 2017 年的 ISLES 挑战赛主要关注从初始 MRI 中预测中风病灶，包括扩散和灌注成像 [70]。与人类读片者之间 DSC 为 0.58 的协议相比，表现最好的模型使用编码器-解码器 CNN，达到了 DSC 为 0.32 [70]。基于该挑战赛的数据，Pinto 等 [71] 提出了结合了 2D 门控循环单元 (GRU) 层的编码器-解码器 CNN，末尾融合了 TICI 评分，根据不同的 TICI 评分生成病灶预测，该模型的 DSC 为 0.35。Nielsen 等 [73] 使用 CNN 基于基线 DWI 和 MR 灌注预测最终的中风病灶，报告了 ROC AUC 为 0.88。研究还发现，训练有治疗或无治疗的数据的 CNN 预测了不同的中风病灶，表明这种模型在探索治疗差异性结果中有潜力。Ho 等 [74] 提出了一种 CNN 模型，直接从 PWI 源图像（即不经过后处理软件生成的参数图）预测病灶，达到相似的 ROC AUC 为 0.871。Yu 等 [75] 展示了带有注意力门控的 U-Net 模型可以在基线 MR 灌注和扩散图像的基础上预测 2-7 天的最终中风病灶，无论再灌注状态如何，获得了中位 DSC 为 0.53 和 ROC AUC 为 0.92。在一项旨在提供更精确的半暗带和缺血核心信息的独立研究中，Yu 等 [76] 通过在部分再灌注或再灌注状态未知的患者中使用 DWI 和 MR 灌注图预训练带有注意力门控的 U-Net 模型，然后用极少再灌注者微调模型以预测半暗带，并用大量再灌注者预测缺血核心。该模型在半暗带上达到了中位 DSC 为 0.60，在缺血核心上达到了 0.57，超越了最先进软件的自动分割性能。采用略有不同的方法，Wang 等 [77] 使用 CNN 在无对比剂的动脉自旋标记 (ASL) 上识别半暗带组织（由对比 PWI 的 Tmax 灌注参数定义），ROC AUC 为 0.958，在 92% 的病例中提供了相似的中风分级，而无需注射对比剂。

从 CT 图像预测最终中风病灶更具挑战性，因为这些标记与组织损伤的相关性不如 DWI。Robben 等 [78] 提出了一个 CNN 模型，该模型从源 CT 灌注图像和临床元数据中进行并行输入，达到了平均 DSC 为 0.48。研究还进行了一项消融研究，显示除了图像信息外，从成像到治疗的时间也影响了模型预测。Amador 等 [79] 应用时间卷积神经网络 (temporal CNN) 来预测基线 CT 灌注源图像的最终病灶，达到了 DSC 为 0.33。Kuang 等 [80] 从 67 位患者的 CT 灌注图和临床数据训练了随机森林模型，并在 137 位患者中进行了测试。研究发现，该模型达到了中位体积差为 -3.2 mL 和 DSC 为 0.388，虽然这些患者的再灌注状态不一致，但该模型显著比阈值方法 (Tmax 阈值和 CBF 阈值) 更准确。

2.5 预测出血性转化

出血性转化是中风治疗的一种潜在并发症，大面积的出血性转化可能会致命。过去，研究人员使用统计方法来预测再灌注治疗后的出血性转化。为了提高预测效果，Yu 等人 [81, 82] 提出了一种长短期记忆网络 (LSTM)，该网络使用基线 MR 灌注图像作为输入，在中风发生 24 小时后通过 GRE 序列识别出血性转化病灶的分割。该模型的 ROC AUC 达到了 0.894，优于先前的支持向量机 (SVM) 方法 (ROC AUC 为 0.837)。Jiang 等人 [83] 结合多参数 MRI 和临床数据来预测出血性转化的发生，将影像序列分别输入到 Inception V3 网络架构中，并在全连接层与临床数据相连接。该模型在出血性转化的二分类任务中取得了较高的 AUC 为 0.932，准确率达到了 0.873

2.6 预测中风临床结局

与在图像上预测未来中风病灶相比，预测临床结局更加困难，原因有几个。最常用的评分系统，即改良 Rankin 评分 (mRS)，具有非线性和主观性，并且分析单位是每位患者而非每个个体素（见表 1）。大多数以往的研究使用非影像数据作为输入，利用简单的统计或更复杂的机器学习模型来预测临床结局 [84-89]。然而，影像可能提供更多信息，例如梗死和出血的空间定位以及脑萎缩的存在。Osama 等人 [90] 提出了一种平行多参数特征嵌入的 Siamese 神经网络 [91]，使用来自 ISLES 2017 挑战赛的 MRI 灌注图和临床数据将 3 个月 mRS 从 0 到 4 进行分类。该模型在每个类别上使用留一交叉验证测试中达到平均准确率为 37%。Nishi 等人 [92] 提出了一种以 DWI 为输入、以中风病灶分割为输出的 U-Net。然后提取 U-Net 的瓶颈特征以预测 3 个月 mRS 是否大于 2，这是良好临床结局的常用指标。该方法的 ROC AUC 达到了 0.81，超过了 ASPECTS 评分 (ROC AUC 为 0.63) 和缺血核心体积模型 (ROC AUC 为 0.64) 的表现。这些研究表明，自动化影像分析在预测临床结局方面可能有帮助，但仍需对这些复杂且具有挑战性的预测进行进一步研究。

改良 Rankin 评分 (mRS)	描述
0	完全无症状
1	尽管有症状，但无明显残疾；能够完成所有日常职责和活动
2	轻度残疾；无法完成所有之前的活动，但能独立处理自身事务
3	中度残疾；需要一定帮助，但能独立行走
4	中重度残疾；无法独立行走，且无法独立处理个人日常需要
5	严重残疾；卧床、失禁，需要持续的护理和照顾
6	死亡

表 1：改良 Rankin 评分 (mRS) 描述

2.7 预测脑血流量 (CBF) 和脑血管储备 (CVR)

有时候，获得更精确的标志物图像（如脑血流量 CBF）对于评估中风严重程度是有用的。当前的 CBF 金标准是 O-15 水正电子发射断层扫描 (PET)，但由于其需要在设施内生产放射性示踪剂并伴有辐射暴露，因而比 MRI 或 CT 的可获得性要低。而动脉自旋标记 (ASL) 是一种非侵入性 MRI 序列，可在不使用静脉注射造影剂的情况下测量 CBF，使得重复检查更为可行，并减少了因造影剂或放射性示踪剂导致的潜在不良影响。尽管过去几十年 ASL 已取得了进展，但其敏感性仍较低，在侧支循环延迟区域经常低估 CBF，并且容易受到多种伪影的影响。

Guo 等人研究了 U-Net 卷积神经网络 (CNN) 是否可以从 ASL 和结构图像生成类似 PET 的 CBF 图 [93]。与 ASL CBF 相比，由 ASL 和结构 MRI 扫描生成的合成 PET CBF 图在结构相似性指数上显著更高 (0.854 ± 0.036 vs. 0.743 ± 0.045)。通过在正常受试者和脑血管疾病患者上进行训练，他们展示了预测 PET CBF 图像的良好性能，无论疾病状态如何。

CVR（脑血管储备）通过计算使用血管扩张药物前后的相对 CBF 变化 ($r\Delta CBF$) 来测量。低 CVR 的患者未来中风的风险更高，识别这些患者可能有助于启动预防性治疗，例如积极的药物治疗、颈动脉内膜切除术或颈动脉支架置入 [93]。乙酰唑胺是一种碳酸酐酶抑制剂，通常用作血管扩张剂来测量 CVR。它通常是安全的，但在对磺胺类药物过敏或患有严重肾脏和肝脏疾病的患者中禁用。某些患者在测试过程中可能会出现类似中风的症状。尽管这些症状是短暂且罕见的，但可能会影响患者和医务人员感到不安。

为了进一步简化 CVR 的测量，Chen 等人 [95] 研究了使用 U-Net CNN 进行无药物 CVR 测量的可行性，这项研究基于 Guo 等人的工作 [93]。该研究还探讨了几种输入组合 (MRI + PET 与仅 MRI)，以确定是否需要基线 O-15 PET CBF 信息。使用 O-15 PET $r\Delta CBF$ 作为真值，在一组烟雾病患者（这种疾病表现为脑动脉慢性狭窄，导致中风风险增加）中，他们表明仅使用基线 MRI 在预测 CVR 受损区域方面的表现优于当前临床方法，即使用乙酰唑胺注射前后的 ASL。此方法可能在临床实践中作为常规 MRI 扫描的一部分来估计 CVR，从而无需 PET 或乙酰唑胺。

3 出血性中风（卒中）或颅内出血

出血性中风，也称为颅内出血，占所有中风的约 13%。尽管缺血性中风的发生率是出血性中风的两倍，但研究发现出血性中风在总死亡人数（每年三百万）和致残率（6900 万伤残调整生命年）上与缺血性中风相似 [96]。出血性中风通常通过非增强 CT 或 MRI 诊断 (GRE 或 SWI 对出血特别敏感)。诊断和分诊的重要考虑因素包括出血的存在、位置、体积和扩张。Chilamkurthy 等人 [97] 用 30 万份 CT 扫描的大规模数据集训练了一个 ResNet，以检测 CT 上的关键发现，包括出血。该模型在 500 例 CT 扫描中进行了测试，检测出血的 AUC 较高。然而，其性能不如专家放射科医生。Lee 等人 [98] 提出了一个 ImageNet 预训练的深度 CNN，并在 904 例急性颅内出血的 CT 病例中进一步训练，以检测出血并分类出血的 5 种亚型。他们在约 400 例的独立测试数据集中进行测试，发现该模型在灵敏度为 92-98% 和特异性为 95% 的情况下，达到了与专家放射科医生相似的性能。此外，研究人员尝试使用注意力图解释此 CNN 模型，显示该模型具有类似于模仿放射科医生工作流程的过程。Kuo 等人 [99] 训练了一个 CNN，使用超过 4000 份头部 CT 扫描来分类和分割颅内出血。他们展示该模型在 200 例独立测试集上获得了 0.991 的 AUC，在非常小和微妙的出血病灶病例中表现良好。

机器学习也被应用于诊断颅内出血的病因，例如微出血、血管畸形和颅内动脉瘤。这些主题在单独的章节中进行了回顾。

4 脑血管畸形

脑血管畸形在普通人群中发生率为 0.1%-4.0%。动静脉畸形（AVMs）是最危险的脑血管畸形，可能导致出血、癫痫发作、头痛和局灶性神经功能缺损。

在非增强 CT 上识别由 AVMs 引起的脑实质内出血可以帮助对患者进行适当的治疗分诊。Zhang 等人 [100] 从 11 种基于滤波的特征选择方法中选择了放射组学特征，并应用多种监督机器学习算法将脑实质内出血分类为与 AVM 相关或其他病因。最佳模型是 AdaBoost 分类器，在测试集上获得了 0.957 的 AUC、88.9% 的灵敏度和 93.7% 的特异性。

立体定向放射外科（SRS）在治疗小于 3 厘米直径的 AVMs 或位于深部和重要功能区的 AVMs 时最为成功，因为这些部位在尝试切除时会带来很大的神经风险。其性能依赖于精确描绘目标 AVM 的准确性，因为部分体积照射可能导致消融失败和残留症状。最近，Wang 等人 [101] 提出了一种三维 V-Net 模型，自动分割对比 CT 图像上的 AVMs，以指导立体定向放射外科。他们将 V-Net 模型的性能与人类读者进行了比较，在 80 名患者中实现了 0.85 的平均 DSC 和 0.076 mL 的平均体积误差。

立体定向放射外科后的不良辐射效应包括可能需要外科干预的囊肿形成和可能导致 1%-3% 患者永久神经功能缺损的辐射引起的变化。深部 AVMs（位于丘脑、基底节和脑干）、大型 AVMs、大辐射治疗体积和重复放射外科是导致放射外科后出现神经功能缺损的风险因素。Lee 等人 [102] 提出了一种无监督分类方法，利用模糊 C 均值聚类分析 T2 加权 MRI 上的 AVM 灶，并分析了邻近灶的脑实质成分与辐射引起的变化之间的关联。该模型在辐射暴露区域自动分割了 AVM 灶、脑实质和脑脊液成分。与手动分割相比，该算法的 DSC 达到了 0.795。自动分割的脑实质与辐射引起的变化相关。

5 颅内动脉瘤

颅内动脉瘤（IAs）在普通人群中的发生率为 3.2%[103, 104]。动脉瘤破裂占自发性蛛网膜下腔出血的 80%-90%[5, 105]，这通常是一个灾难性的事件，死亡率为 23%-51%[4, 5]，并且 30%-40%[4, 6] 的患者会出现永久性残疾。幸存者常常面临长期的神经心理缺陷和生活质量下降。尽管数字减影血管造影（DSA）是诊断动脉瘤的金标准，但未破裂的颅内动脉瘤可以通过非侵入性成像技术如磁共振血管成像（MRA）或 CT 血管成像（CTA）来检测。早期诊断颅内动脉瘤有助于临床管理，可以预防其破裂 [106, 107]。然而，颅内动脉瘤的诊断和管理仍存在两个未满足的临床需求。

5.1 动脉瘤检测中的困难

由于颅内动脉瘤 (IAs) 体积小以及颅内血管的复杂性，动脉瘤的检测可能耗时且需要专业培训。这带来了两个挑战。首先，在从 CT 血管造影 (CTA) 和 MRI 血管造影 (MRA) 中检测颅内动脉瘤时，观察者之间的协议较差 ($\kappa = 0.67-0.73$) [108]。解读结果可能因专业水平而异。因此，在 CTA 和 MRA 中检测颅内动脉瘤的灵敏度可能在 60% (住院医生) 到 80% (神经放射科医师) 之间 [109]。其次，在检测直径小于 5 毫米的小动脉瘤时，假阴性率较高。报道显示，CTA 检测小于 5 毫米的颅内动脉瘤的灵敏度为 57%-70% [108, 110]，而 MRA 为 35%-58% [109, 110]。相比之下，检测大于 5 毫米的颅内动脉瘤的灵敏度分别为 94% (CTA) 和 86% (MRA)。鉴于上述所有困难，临幊上需要高性能的计算机辅助诊断 (CAD) 工具来帮助检测，提高效率，减少观察者之间的分歧，从而可能改善患者的临幊护理。

5.1.1 颅内动脉瘤检测的人工智能算法

已有多项研究表明，计算机辅助诊断 (CAD) 程序可以自动检测 MRA 或 CTA 中的颅内动脉瘤 (IA)。基于手动设计成像特征（如血管曲率、阈值处理或区域生长算法）的传统 CAD 系统在检测 IA 方面表现良好 [111, 112]。然而，这些传统方法是在非常小的数据集上开发的，并且在应用于新图像时必须手动进行修改。新的基于深度学习的方法可以直接从大量标记图像的数据集中学习最具预测性的特征。与传统方法相比，它们具有更好的性能和更强的泛化能力。深度学习也已被用于 MRA 和 CTA 中的 IA 检测，几项研究显示了不错的结果 [113-116]。

不同成像方式的模型诊断准确性已经得到研究。数字减影血管造影 (DSA) 是一种侵入性血管成像程序，是诊断动脉瘤的金标准。Zeng 等人 [117] 通过将 DSA 图像块的五个连续旋转角度拼接作为模型输入，应用了 2D 卷积神经网络 (CNN) 于 3D DSA。该模型达到了 99% 的准确率。Duan 等人 [118] 进行了类似的任务，但使用的是 2D DSA。这一任务更加困难，因为在 2D 投影图像中可识别的特征较少，特别是在区分血管重叠和动脉瘤时。他们提出了一个两阶段的检测系统：首先，神经网络使用特征金字塔网络在 DSA 上定位目标区域。其次，通过将前后视图和侧视图的双输入传递给另一个特征金字塔网络，生成动脉瘤和血管重叠的锚框。该模型达到了 93.5% 的 AUC。

MRA 和 CTA 提供了颅内动脉瘤的非侵入性诊断。Nakao 等人 [113] 和 Sichtermann 等人 [119] 展示了使用卷积神经网络 (CNN) 进行时间飞行 (TOF) MRA 上动脉瘤检测的可行性。最近，Ueda 等人 [114] 训练了一个 ResNet-18 模型，用于检测 683 个 TOF MRA 上的动脉瘤。该模型在内部数据和外部数据上的测试结果显示，灵敏度和特异性均超过 90%。Park 等人 [115] 提出了一种具有编码器-解码器结构的 3D CNN，从 CT 血管造影中分割颅内动脉瘤。与 U-Net 类似，该模型包含跳跃连接，以便直接将输出从编码器传输到解码器。该编码器使用标记有人类动作的视频进行预训练。模型使用 611、92 和 115 个 CTA 进行了训练、验证和测试。与未增强相比，使用人工智能生成的分割结果对医生的辅助提高了灵敏度、准确率和观察者间一致性。Faron 等人 [120] 在较小数

据集中也展示了在 3D TOF MRA 中的类似结果。

5.2 动脉瘤风险评估的困难

一旦在影像学检查中发现颅内动脉瘤 (IA)，临床医生必须决定如何处理未破裂的动脉瘤。总体而言，IA 的年破裂风险较低，为 0.95%[121]。当前预防 IA 破裂的治疗方法包括开放神经外科夹闭或血管内栓塞；这两种方法均具有相对较高的围手术期中风和死亡风险 (3%-10%) [122]。因此，未破裂动脉瘤的管理仍然存在争议 [123]。目前，是否进行干预的决定主要基于动脉瘤的大小。如果 IA 在前循环中直径大于 5 毫米，或在后循环中大于 7 毫米，则考虑进行外科治疗 [123]。如果 IA 小于这些阈值，通常会进行随访观察，采用序列影像学检查 [124]。在随访期间，IA 的大小变化是即将破裂的警告信号，通常会导致外科或血管内治疗。然而，IA 的破裂除了大小外，还依赖于多个因素，包括动脉瘤的形状和位置、动脉瘤的血流动力学、血压以及患者的心理和身体压力 [121, 125]。仅依据大小标准做出干预决定并不理想，因为破裂风险是多因素的。此外，随访序列影像学检查需要时间，破裂可能在观察期间发生 [126-128]。

5.2.1 基于人工智能的动脉瘤风险预测模型

对颅内动脉瘤 (IA) 进行更全面的形态评估是理想的；这应包括动脉瘤的形状、几何特征、是否存在分支囊、体积以及对比序列扫描中的 IA 形态。基于深度学习的方法有潜力自动执行精确的 IA 分割，并为 IA 的形态评估提供有效工具。此外，机器学习方法可以处理高维的跨领域输入，并直接从标记数据中学习，以构建复杂的预测模型。机器学习模型派生的特征排名可以提供关于影响模型预测的个体因素的信息。

已有几项研究尝试使用深度学习对动脉瘤进行分割 [129, 130]。Podgorsak 等人 [130] 使用具有编码器和解码器架构的卷积神经网络 (CNN) 对 DSA 上的动脉瘤进行分割，达到了颅内动脉瘤的 DSC 值超过 0.9。需要优化对未破裂小动脉瘤 [以及多重动脉瘤患者] 的治疗决策。研究已应用机器学习算法预测未破裂动脉瘤的预后 [131-137]。Liu 等人 [132] 利用从 DSA 派生的形态特征和机器学习模型预测动脉瘤是否不稳定（定义为在 1 个月内破裂）、动脉瘤的生长及症状性动脉瘤。他们发现，直径在 4 到 8 毫米之间且形态不规则的动脉瘤表明动脉瘤不稳定，在一个独立的测试集中，其曲线下面积 (AUC) 为 0.85。类似地，Kim 等人 [133] 在小动脉瘤的旋转 DSA 图像上使用 CNN，结果显示该模型在动脉瘤破裂预测方面的性能优于人类预测。

Tanioka 等人使用基于机器学习的方法，结合形态学和血流动力学参数作为输入，成功预测 IA 的破裂状态，准确率相对较高 (71.2%-78.3%) [138]。他们发现，投影比、形态不规则性和大小比对破裂动脉瘤的鉴别具有重要意义。Shi 等人进一步将临床数据纳入形态和血流动力学信息，以构建机器学习模型来预测 IA 破裂，并报告曲线下面积为 0.88-0.91[139]。

在动脉瘤破裂后，预测动脉瘤性蛛网膜下腔出血的常见并发症，如血管痉挛、延迟性脑缺血和功能结果，可以帮助指导患者护理。Kim 等人 [140] 使用临床因素和动脉瘤的形态特征，通过随机森林回归器预测 IA 破裂后的血管痉挛。该模型的准确率达到 0.855 (AUC 为 0.88)。Ramos 等人 [141] 使用临床和 CT 图像特征，应用多种机器学习算法预测延迟性脑缺血。最佳模型的 AUC 达到 0.74。类似地，Rubbert 等人 [142] 使用临床和影像特征，通过随机森林预测 6 个月的修订 Rankin 量表分级，准确率为 71%。

6 脑小血管病

脑小血管病 (cSVD) 涵盖了一系列影响大脑小穿透动脉、毛细血管以及可能的静脉的小血管疾病 [143]，这些疾病引起各种局部和全脑病变，可通过病理检查和脑成像检测到 [144]。cSVD 具有广泛的临床表现。尽管许多受影响的患者可能保持无症状，但 cSVD 可能预示着患者面临急性缺血性中风或脑内出血的风险；它也可以表现为与逐渐认知下降、情绪障碍和步态紊乱相关的潜隐性临床过程 [145]。cSVD 导致约四分之一的急性缺血性中风，并且是出血性中风的主要风险因素 [146-148]。它是血管性痴呆和混合性痴呆最常见的原因，通常与阿尔茨海默病同时发生，且大约造成全球一半的痴呆，因此造成巨大的健康负担 [146, 149, 150]。

6.1 脑小血管病的影像学特征

神经影像学在脑小血管病 (cSVD) 的诊断和评估中发挥着关键作用 [143]。根据“神经影像学血管变化报告标准”(STRIVE)，cSVD 的影像学特征包括近期的小皮层下 infarcts、推测为血管来源的白质高信号 (WMH)、腔隙、扩大的血管周围间隙 (PVS) 和脑微出血 (CMBs) (图 4) [144]。这些影像学发现，无论是单独还是结合在一起，都与认知障碍、痴呆、抑郁、运动问题、中风风险增加以及中风后的不良结局相关 [146, 151-153]。对 cSVD 影像学特征的量化对于疾病严重性评估和临床预后判断非常重要 [154, 155]。然而，这些病变通常较小且在大脑中广泛分布，使得手动检查和分割工作繁重且容易出错。机器学习算法在 cSVD 影像学特征的自动量化方面具有巨大潜力。通过结合所有相关特征，可以计算出大脑的“总 cSVD 评分”，这可能更好地反映 cSVD 的疾病状态和负担。这些应用可以帮助对 cSVD 患者进行疾病诊断、治疗、监测和预后评估。我们将回顾当前机器学习在 cSVD 影像学特征检测和量化中的应用，包括 WMH、CMB、腔隙和 PVS，以及 cSVD 的总负担。

	Recent small subcortical infarct	White matter hyperintensity	Lacune	Perivascular Space	Cerebral microbleed
Clinical Image					
Illustration					
Diameter	$\leq 20 \text{ mm}$	Variable	$3\text{-}15 \text{ mm}$	$\leq 2 \text{ mm}$	$\leq 10 \text{ mm}$
DWI	↑	↔	↔/(↓)	↔	↔
FLAIR	↑	↑	↓	↓	↔
T2	↑	↑	↑	↑	↔
T1	↓	↓	↓	↓	↔
T2*/SWI	↔	↑	↔	↔	↓↓

图 4: 脑小血管病的 MR 成像特征。(上部) 临床图像 (上部) 和脑小血管病 MRI 特征的插图 (中部), 以及各特征影像学特征的总结 (下部)。DWI 指扩散加权成像, FLAIR 指液体衰减反转恢复, SWI 指磁敏感加权成像。其中, ↑ 表示信号增强, ↓ 表示信号减弱, ↔ 表示等信号 (此图根据 Wardlaw 等人的研究 [145] 改编)。

6.2 白质高信号分割

推测为血管来源的白质高信号 (WMH), 在流体衰减反转恢复 (FLAIR) MRI 中表现为白质内的高信号病灶, 是脑小血管病 (cSVD) 的主要特征之一 [144]。这些异常在正常衰老、痴呆和中风中起着关键作用。大规模纵向人群研究证实了 WMH 容积与临床结果之间的剂量依赖关系, 使其测量具有临床价值 [156]。Fazekas 视觉评分量表是临床环境中评估 WMH 负担的最广泛使用的方法; 该量表分为四级, 根据侧脑室周围和深白质区域 WMH 病灶的大小和连通性进行评分 (图 5) [157]。然而, Fazekas 量表具有较高的个体内和个体间变异性 [158], 显著的天花板 / 地板效应 [159], 以及对临床组差异的敏感性较差 [160], 导致 WMH 研究中的不一致性。

需要进行 WMH 病灶容积的分割和量化。在深度学习技术出现之前, 提出了许多自动 WMH 分割方法, 包括监督方法, 例如 k 近邻法 [161]、支持向量机 [162]、基于信号强度和空间信息的贝叶斯方法 [162] 或多对比图像 [163]、结合形态分割和自适应提升分类器 [164], 以及人工神经网络 [165], 还有无监督方法, 例如直方图分析 [166]、模糊分类算法 [167, 168]、高斯混合模型 [169] 和隐马尔可夫随机场模型 [170]。然而, 这些方法通常限于特定成像方式和患者特征 (例如年龄、临床表现), 并使用不同的分析指标, 使得方法之间的比较变得困难 [171]。

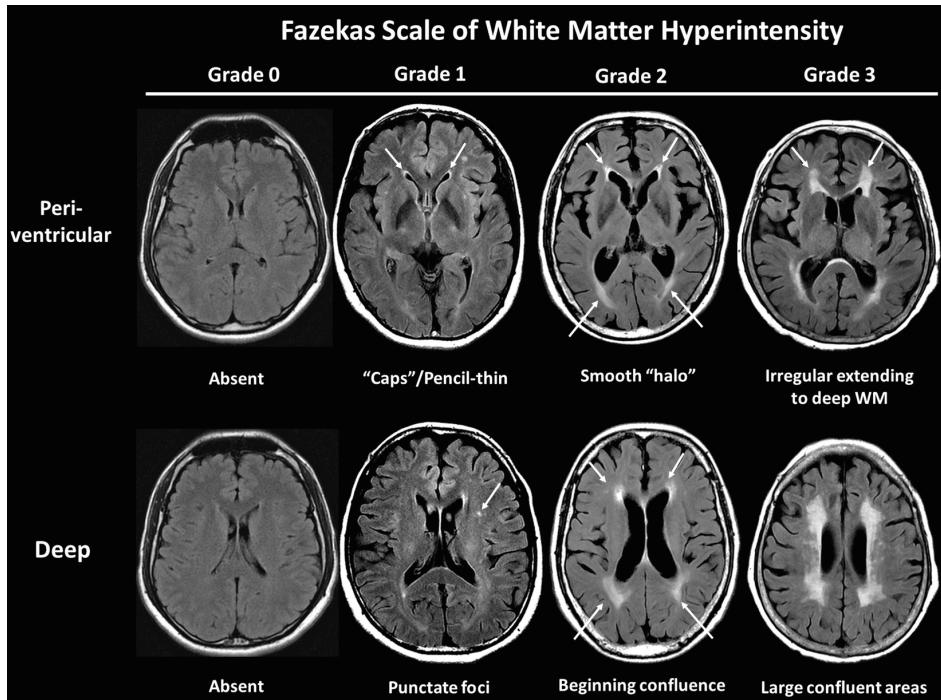


图 5: Fazekas 视觉评分量表用于评估白质高信号。该量表根据病灶的大小和融合程度分为四级，分别针对脑室周围（上部）和深层白质（下部）区域。

6.2.1 基于深度学习的白质高信号分割方法

在 2017 年的医学图像计算与计算机辅助干预大会 (MICCAI) 上，开展了白质高信号 (WMH) 分割挑战赛 (<https://wmh.isi.uu.nl/>)，为 WMH 分割的自动化方法提供了标准化评估。该多中心/多扫描仪数据集包含了不同程度的与年龄相关的退行性和血管病理的患者图像。训练数据集包括来自 3 台扫描仪的 60 幅图像，由 2 位专家手动分割 WMH 作为基础真相。测试数据集则包括从 5 台 MR 扫描仪获取的 110 幅图像，其中包括来自 2 台未在训练集中使用的扫描仪的数据，以评估分割方法在未测试扫描仪上的可推广性。使用了五个评估指标（包括 DSC、修改的 Hausdorff 距离、体积差异、灵敏度和 F1 分数）来对方法进行排名。在 20 个参与者中，前 10 名参与者均应用了深度学习方法 [172]。排名靠前的方法的表现与两位独立的人工观察者相似或更好，后者并未作为基础真相的评定者，这表明自动化方法替代人工评定者的潜力（图 6）。Li 等人 [173] 获得了第一名，通过利用与 U-Net 相似的三种完全卷积神经网络的集成，达到了 0.8 的 DSC 和 0.84 的召回率。值得注意的是，他们在前 1/8 和最后 1/8 切片中去除了 WMH 预测，因为假阳性预测经常发生，这是一种后处理方法。Andermatt 等人 [174] 获得第二名，利用基于多维门控递归单元 (GRU) 的网络，在 3D 补丁上进行训练，达到了 0.78 的 DSC 和 0.83 的召回率。Ghafoorian 等人 [175] 获得第三名，构建了一个多尺度 2D CNN，在十折交叉验证中训练，并选择训练数据中表现最好的三个检查点，达到了 0.77 的 DSC、0.73 的召回率和最高的 F1 分数 0.78。Valverde 等人 [176] 获得第四名，构建了三个 3D CNN 的级联框架，第一个模型用于识别候选病灶体素，第二个减少假阳性

检测，第三个执行最终的 WMH 分割。总体而言，挑战赛的结果表明，集成方法和减少假阳性的策略（包括选择性采样 WMH 模拟、去除容易出现假阳性的切片以及添加假阳性减少模型）是有利的。排名靠前的模型在 FLAIR 上呈高信号但非 WMH（例如，透明隔膜）正常区域的假阳性极少，这是许多低排名方法的缺陷。尽管排名前四的模型在扫描仪间的鲁棒性排名中仍然是领先者，但一些高排名的基于深度学习的方法在扫描仪间鲁棒性方面表现不如低排名的规则基础方法，这表明数据驱动的方法有时可能不适合未见过的扫描仪。

WMH 分割挑战赛仍然接受新的和更新的提交。Zhang 等人 [177] 设计了一种双路径 U-Net 分割模型，该模型利用注意力机制结合 FLAIR 序列和大脑图谱（用于位置信息）输入，实现了比之前提到的方法更高的性能。Park 等人 [178] 提出了一种具有多尺度突出前景的 U-Net，旨在改善具有部分体积效应的 WMH 体素的检测，达到了创纪录的 DSC (0.81) 和 F1 分数 (0.79)。

尽管深度学习方法越来越受欢迎，并在 WMH 分割挑战中表现出色，但一项近期的系统评价 [179] 显示，2015 年至 2020 年 7 月间开发的自动 WMH 分割方法并未提供证据支持深度学习方法在临床研究中优于 k-NN 算法 [180, 181]，线性回归 [182, 183] 或无监督方法（如模糊 C 均值算法 [184, 185]、高斯混合模型 [186]、统计定义 [187]），在与参考分割的空间一致性（即 DSC）方面。非深度学习方法，如 k-NN 和线性回归方法，具有简单性，可以更容易训练，并且在处理有限数量的训练数据时可能不易过拟合。未来的研究需要高质量的大型开放数据和代码可用性，以克服研究设计和基础真相生成中的偏倚，从而全面比较和验证这些方法 [188]。

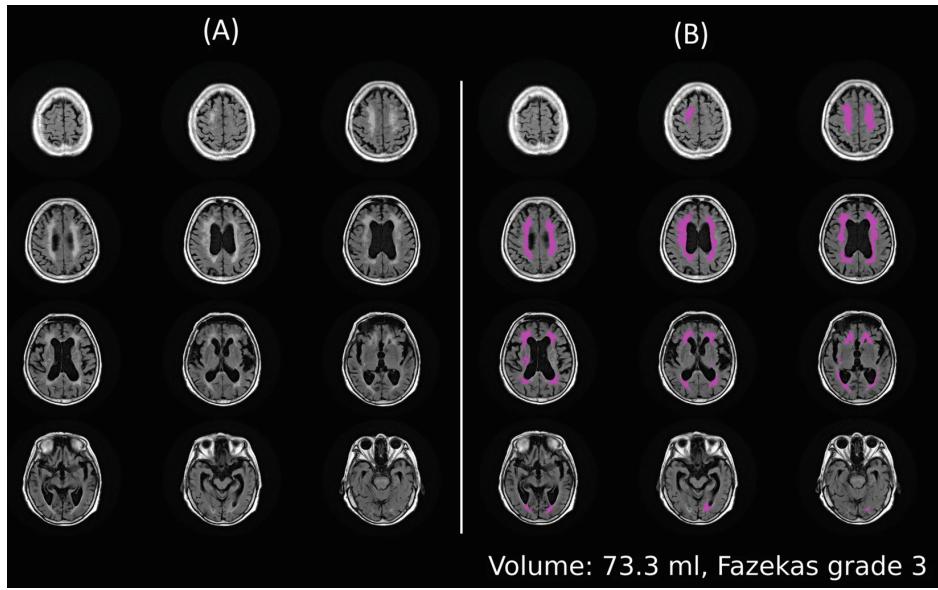


图 6：白质高信号（WMH）分割和量化的示例。（a）原始 T2 FLAIR 图像。（b）通过深度学习算法可以实现自动 WMH 分割（粉色区域）和体积量化，相较于 Fazekas 量表，这种方法能更精确地估计大脑中的 WMH 负担。WMH 指白质高信号。

6.3 脑微出血 (CMB) 检测

脑微出血 (CMB) 是脑小血管病的影像学表现，通常定义为在 T2 加权渐变回波 (GRE) 或敏感加权成像 (SWI) 上显示的直径小于或等于 10 毫米的信号缺失区域。CMB 常见于自发性颅内出血 [189] 或认知障碍 [189] 的患者，并与静脉溶栓或治疗性抗凝后出血风险增加相关 [190, 191]。CMB 与潜在的未控制高血压（特别是在深部及/或后颅窝结构中）[192] 和/或脑淀粉样血管病（特别是在皮层位置观察到时）[193] 高度相关。检测 CMB 在评估中风患者治疗规划中的益处和风险时具有临床重要性。

Greenberg 等人 [189] 发表了一本关于 CMB 检测的详细指南。CMB 的微小尺寸以及多种 CMB 假象（如小静脉、钙化、海绵样畸形、深核中的铁沉积和流 voids）导致观察者之间的一致性有限、扫描解读时间较长，且手动检查的错误率增加，尤其在重度 CMB 负担的患者中。

自动 CMB 检测方法可能会提高 CMB 识别的效率和准确性。目前，已有基于放射组学和传统机器学习的自动检测方法进行研究。Van den Heuvel 等人 [194] 利用 CMB 的暗色和球形特征，并应用随机森林分类器，实现了 89.1% 的敏感性和每个受试者 25.9 个假阳性。几项研究已应用深度学习模型以改善 CMB 检测 [195-198]。Dou 等人 [198] 利用两步级联框架，首先使用 3D 完全卷积网络筛查 CMB 候选者，随后用 3D CNN 鉴别器排除 CMB 假象，达到了 93.16% 的敏感性、44.31% 的准确率和每个受试者 2.74 个假阳性。Liu 等人 [196] 采用两阶段 3D CNN 架构，并将相位图像添加到 SWI 中作为模型输入。相位图像能够区分顺磁性 CMB 和抗磁性钙化，而这一点仅通过 SWI 是无法做到的。他们的模型成功降低了假阳性检测，并达到了 95.8% 的敏感性、70.9% 的准确率和每个受者 1.6 个假阳性。Rashid 等人进一步将定量敏感性成像 (QSM) 添加到 SWI 中作为输入，构建了一个多类 U-Net CNN 方法，以区分 C 和非出血性铁沉积，而这在仅使用 SWI 和相位图像时是无法实现的 [197]。该多类模型在 CMB 检测中达到了 84% 的敏感性和 59% 的准确率，而在铁沉积检测中达到了 75% 的敏感性和 75% 的准确率。

6.4 腔隙性病灶检测

腔隙性病灶通常是由于小型慢性皮质下梗死或出血，位于穿通小动脉供血的深层灰质和白质中所产生的后遗症 [144]。这些病灶与中风、痴呆和步态障碍的风险增加相关 [143, 144]。在神经影像学上，腔隙病灶表现为圆形或卵圆形的皮质下充液腔隙，直径在 3 到 15 毫米之间，通常在 T2 FLAIR 图像上显示出周围的高信号胶质反应圈 [144]。纵向空间映射研究表明，新出现的白质高信号 (WMH) 常出现在小型皮质下梗死周围 [199]，而新的腔隙病灶则出现在 WMH 的边缘 [200]，表明两种病变类型之间存在较强的关联和邻近性。因此，能够自动分割 WMH 并检测腔隙病灶的应用受到重视。然而，目前很少有研究提出腔隙病灶的自动检测方法。Uchiyama 等人 [201] 开发了一种算法，首先使用顶帽变换和多相位二值化技术检测潜在的腔隙病灶候选区域，然后使用基于规则的方案和支持向量机来去除假阳性，达到了 96.8% 的灵敏度和每切片 0.76 个假阳性。Wang 等人 [169] 应用

了一种多步骤算法来检测 WMH、皮质梗死和腔隙病灶，包括脑组织提取、高斯混合模型分割脑组织中的高信号病灶、基于解剖位置和形态学操作将 WMH 与皮质梗死分离，并基于位置和强度阈值分割腔隙病灶。该方法在腔隙病灶检测中达到了 83.3% 的灵敏度和每个受试者 0.06 个假阳性。Ghafoorian 等人 [202] 使用了两阶段的深度学习方法，其中包括用于候选检测的全卷积神经网络和用于减少假阳性的 3D 多尺度位置感知 CNN。该方法在腔隙病灶检测中达到了 97.4% 的灵敏度和每切片 0.13 个假阳性。

6.5 血管周围空间定量

血管周围空间 (PVS)，也称为 Virchow-Robin 空间，是围绕脑内穿透血管的脑外液体空间的延伸 [144]。最近，它们被认为是脑内一种广泛的血管周围液体运输系统的一部分，即淋巴系统，负责清除脑内废物 [203]。正常的 PVS 在传统 MRI 上通常不可见，而增大的 PVS 则与皮质下梗死、白质高信号 (WMH)、皮质微出血 (CMB) 和认知衰退的进展相关，被认为是慢性小血管病 (cSVD) 的生物标志物 [204]。在神经影像学上，PVS 表现为直径小于 3 毫米的圆形或卵圆形腔隙，其信号强度与脑脊液 (CSF) 相同。它们通常位于下基底神经节、中心半卵圆中心和中脑。PVS 在 MRI 上可能与腔隙病灶相似，但 PVS 没有周围的胶质反应圈，并且在平行于穿透血管的方向成像时显示出更长的形状。PVS 的严重程度可根据 Charidimou 等人提出的四分制视觉评分量表进行分级，具体分为 0 级 (无 PVS); 1 级 (轻度，1-10 个 PVS); 2 级 (中度，11-20 个 PVS); 3 级 (中重度，21-40 个 PVS); 4 级 (重度，>40 个 PVS)，适用于基底神经节和中心半卵圆中心 [205]。由于 PVS 尺寸较小且数量众多，手动计数或分割 PVS 极其繁琐且耗时，这可能解释了文献中自动 PVS 定量方法研究的稀缺性。Park 等人 [206] 提出了基于手动衍生的 PVS 掩膜在 7T MR 图像上的自动 PVS 分割方法。他们从由脑和血管结构确定的兴趣区域中提取了常用于对象识别的 Haar-like 特征，并使用随机森林分类，达到了 0.73 的 DSC、69% 的灵敏度和 80% 的阳性预测值。Ballerini 等人 [207] 提出了基于 3D Frangi 滤波的 PVS 分割。由于缺乏 PVS 分割掩膜的标准答案，他们使用有序 logit 模型和视觉评分量表进行了方法的优化和评估，该方法在基于分割的 PVS 负担和视觉评分量表之间达到了 Spearman 相关系数为 0.74 ($p < 0.001$)。Dubost 等人 [208] 使用 3D 卷积神经网络回归预测视觉评分量表，并在视觉和自动评分量表之间达到了 0.75-0.88 的组内相关系数 (ICC)，甚至高于人工评分者之间的评分一致性。

6.6 总小血管病负担

慢性小血管病 (cSVD) 被认为是一种动态的全脑疾病，具有广泛的临床表现和弥散的脑影像学表现，同时共享共同的微血管病理 [209]。结合所有影像学特征的多因素方法可能更好地代表 cSVD 的负担和疾病状态。已经引入了几种 cSVD 总负担的视觉评分系统 [154, 205]。Staals 等人 [154] 提出了一个四分制评分系统，其中每出现一种 cSVD 影像学特征即得一分：(1) 多于一个腔隙性梗

死；(2) 多于一个微出血；(3) 基底神经节中度至重度（多于 11 个）PVS；以及 (4) 脑室周围白质高信号 (WMH) Fazekas 评分为 3 和/或深部 WMH Fazekas 评分为 2-3。尽管这些半定量评分系统在临床应用中是实用且简单的，它们也有一些局限性。首先，它们可能不够敏感，无法代表疾病的严重程度，因为 cSVD 负担的累积形成可能是主观且对评分者来说繁琐的，尤其是在 WMH 和 PVS 的评估中。其次，现有的评分没有考虑病变的位置，但解剖位置是已知的认知损伤关键因素 [210]。前面章节中描述的不同 cSVD 影像特征的自动方法可以提供全脑 cSVD 负担的定量测量，并且非常适合克服这些限制。几项研究表明，计算机生成的 cSVD 总负担在 cSVD 患者评估中具有巨大潜力。Duan 等人 [211] 开发了一个基于多种 CNN 的系统，可以在 4.4 秒内准确分割皮质下梗死、皮质微出血 (CMB)、WMH 和腔隙性病灶。Dickie 等人 [212] 在多对比度 MR 图像上使用基于体素的高斯混合模型聚类分析来估计整体 WMH、腔隙性病灶、CMB 和萎缩为“脑健康指数”；他们表明脑健康指数与认知结果的相关性强于 WMH 体积和视觉 cSVD 评分。Jokinen 等人 [213] 使用基于自动化图谱和 CNN 的分割方法来获得 WMH、腔隙性病灶、PVS、皮质梗死和脑萎缩的体积测量，表明所有标记的组合测量比任何单独的测量更能有效预测认知和功能结果。

总体而言，先前的研究表明，机器学习算法在执行 cSVD 影像特征的自动分割或检测方面具有巨大潜力。通过结合每个 cSVD 特征的测量，可以量化“cSVD 总负担”，这可能用于促进 cSVD 患者的临床评估、治疗监测和结果预测（如图 7 所示）。

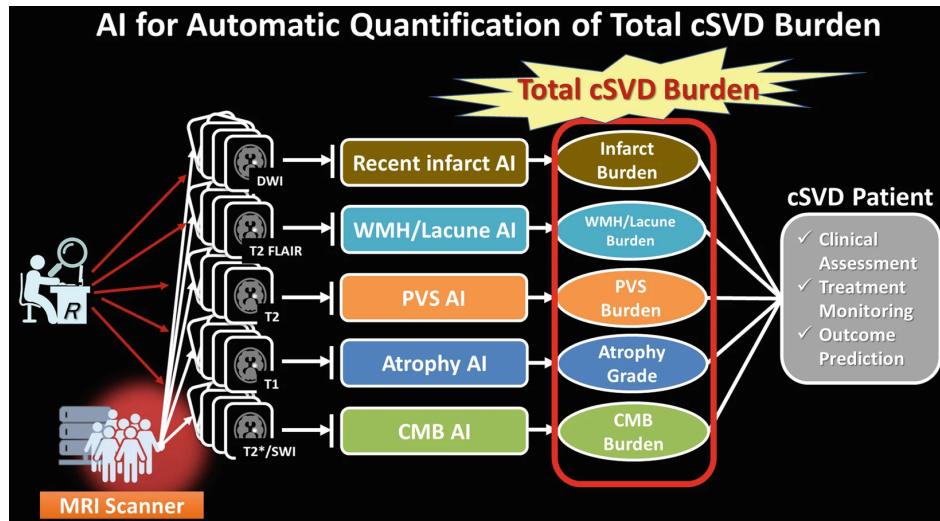


图 7：脑小血管病 (cSVD) 的 AI 应用。AI 算法在对 cSVD 影像特征进行自动量化方面具有巨大潜力。通过结合这些负担，可以量化“总 cSVD 负担”，这可能有助于 cSVD 患者的临床评估、治疗监测和结果预测。

7 结论

总之，机器学习算法在改善脑血管疾病的临床诊断和护理方面展现出巨大潜力。尽管机器学习的性能因研究和数据集而异，但在许多情况下，已超过当前临床的最佳水平。亟需更多的大型队列验证

研究，同时开发标准测试集以比较不同算法，这将有助于实现更公正的比较。此外，还需要更多的实际应用经验，以理解机器学习在改善脑血管疾病的诊断和护理中的作用。

参考文献

[1-213]