

P898(Challenges in the Path 之后)-P903;P905-P916;P919-P933;P939-P947;P951-P957(BE Cancer Detection Using Narrow-Band Imaging 之前)

P898 (Challenges in the Path 之后)-P903;

研究上的挑战

1. 来自注释的挑战

如前一节所述，人工智能系统需要外科医生不断进行注释：

时间限制是最重要的挑战，因为外科医生的工作日程经常很紧，这需要他们全神贯注。

另外，手术数据庞大，注释需要专家知识，因此，这是一项昂贵且耗时的活动。

由于需要来自多家医院的数据来提供足够的信息来代表 AI 系统将要学习的任务，因此患者保密和隐私面临更大的挑战。

此外，目前还没有任何标准化的数据采集和注释政策，也没有数字格式的最小数据。

应对这一挑战的建议：

主动学习：从大量数据集中选择重要数据点进行注释 [16] 。

众包：一种解决问题的方法，其中一项任务被委托给一个个人团队，而不是一个人或几个人。大型团队可能包含相对新手，但他们的经验不足可以通过团队成员分散的智慧来弥补。因此，这种方法在灵活性、效率和可扩展性方面是有利的。众包已成功应用于结肠和息肉的诊断以及疟疾感染红细胞的鉴定等医疗保健领域，在颌面外科领域也有前景 [17] 。

本体论的方法：帮助保持外科医生用于注释的词汇的一致性，使其易于被 AI 系统读取。

2. 政策挑战

监管和报销计划将决定现场人工智能的获取和可用性。

应对这一挑战的建议：

在这个方向上，外科医生没有多少可以直接做的。然而，政策制定者肯定会将外科医生的意见作为制定指南的优先考虑因素。因此，保持详细的标准操作程序流程图和人工智能在手术中的优势注释，并有证据支持，将证明是有益的。

3. 算法复杂性和偏见挑战

人工智能将在一段时间内成为医学专家的一个新课题。人工智能系统的实施不受医务人员的直接管辖。医生/外科医生的主要职责也不是了解人工智能系统的细微差别和功能。因此，企业或其他掌权机构有可能通过滥用人工智能算法功能的复杂性来掩饰错误的意图，如歧视。这种情况的一个例子是非营利组织 Data & Society 的 Tim Hwang 和 Madeleine Clare Elish 对网约车公司 Uber 提出的索赔。指控称，优步使用激增定价算法作为前置词，通过向终端用户描绘需求的虚假外观来实施不公平定价[18]。

应对这一挑战的建议：

任命一个由不同专业人士组成的董事会或委员会，以避免既得利益和滥用可能的算法偏见。

了解人工智能的基本原理，以警惕不当行为。

确保严格的法律蓝图和政府法律，以根除人工智能恶意牟利的不道德杠杆。

文献综述：人工智能目前在颌面部外科手术控制的潜在应用

颌面外科前成像

现状

人工智能正被用于图像采集和解释，以增强手术中的诊断、预测和风险分析。数字射线照相术已经取代了传统方法，将人工智能纳入口腔内扫描设备，除了人工智能数据采集之外，还带来了一个额外的、急需的直接优势，即减少患者的辐射暴露。AI 通过优化信噪比确保低辐射剂量的高分辨率图像 [19] 。

借助“无线电显微技术”量化肿瘤表型和决策支持，这是一种人工智能方法，可以分析大量的射线图像，然后从中提取相关的定量特征 [20] 。

潜在的

深度学习技术可用于从若干重复的常规检查中的特定图像的标准分层表示中学习，然后区分正常和异常的射线照相发现。

卷积神经网络与深度学习一起具有从放射图像中识别疾病特征特征以预测病变/疾病实体性质的潜力。

重建和叠加多种诊断模式，如患者口腔内图像、数字照片和 CBCT。开发雄心勃勃的 AI 系统，集成 3D CBCT 图像、计划中的数字模拟手术截骨术切口和机器人手臂，以实现术中精确截骨术并检查患者在手术台上的动态咬合，从而以最短的住院时间和恢复时间实现高度优越的面部重建。

种植手术

现状

一项机器人手臂将颧骨植入假体颅骨的初步研究显示，其准确性高于传统的手术植入。作者希望在未来通过添加触觉传感器和其他人工智能工具来增强他们的发现 [24] 。

潜在的

AI 可以诊断骨密度和骨质疏松症，这在临幊上与植幊手术有关。人工智能模型已经使用全景照片来区分骨质疏松和健康受试者。AI 算法可以评估牙槽骨皮质宽度和皮质侵蚀，以 95% 的准确性、敏感性和特异性区分骨质疏松和健康受试者。这些初步研究有利于在未来将此类人工智能模型应用于常规手术规划 [25] 。

颞下颌关节（TMJ）手术

现状

日本研究人员开发了一种咀嚼机器人颅骨，它可以通过电线和电机工作。它采用人工肌肉致动器（AMA）来模拟动态咀嚼肌活动和咀嚼。然而，这是一个更适用于颞下颌关节手术培训和研究的通用系统，很难构建以复制每个患者 [26] 。然而，这无疑可以为未来的肌肉骨骼疾病研究奠定基础，方法是为个体患者制作定制的机器人头骨，并通过 CT 扫描和面部照片/头颈部印记进行测量。这对于管理诸如肌筋膜疼痛不适综合征（MPDS）、触发点肌肉疼痛、纠正由肌肉松弛或低/高活动引起的异常下颌运动等情况尤其有用。由于其多因素

病因、与牙合的关系以及与头颈部颞下颌关节功能的相互关系，这些疾病的诊断通常具有挑战性。

由于闭塞、关节内间隙可视化、颞下颌关节相互关系（右和左）等多个相互关联的因素，颞下颌骨关节的内部错乱一直是一个谜一般的复杂难题。临床医生不得不依赖外部临床评估模式，因此很难做出明确的诊断和最佳治疗。土耳其研究人员利用患者临床症状和 TMJ 内部紊乱的后续诊断等数据来训练人工神经网络。该结果具有良好的诊断效果 [27]。

潜在的

由于靠近腮腺、面神经和颈外动脉分支等重要解剖结构，颞下颌关节的手术解剖和入路很棘手。因此，居民和新手外科医生发现在这个地区进行手术很有挑战性。人工智能和深度学习工具可以作为导航、面部平面识别和神经传感器设备来拯救外科医生。

全同种异体关节置换术是颞下颌关节强直治疗的一种新兴选择。AI 可用于精确制造和放置这些关节，以改善结果，尤其是与目前仅限于旋转运动的同种异体关节的运动范围有关的结果。

肿瘤外科与重建

现状

IBM Watson 是一个基于人工智能的认知系统，它利用自然语言处理和动态学习，在分析患者报告后为临床医生提供基于证据的治疗选择。肿瘤学 IBM Watson 由纪念斯隆凯特琳癌症研究所的专家培训，目前正在全球医院使用。

肿瘤切除后颌骨的重建具有挑战性，要记住头部和颈部的美学和功能。CTGAN 是一种深度卷积生成对抗网络（DCGAN），已尝试在根治性肿瘤手术后确定正常易患颌骨形态，并取得了良好的结果[28]。

潜在的

众所周知，肿瘤手术是一种耗时的手术，需要患者全身麻醉数小时。如早期随机对照试验所示，用于有效管理患者生命体征（如血压）的人工智能系统将为手术麻醉团队提供极大的缓解 [29]。

智能外科医生的手术刀/刀集成了 AI 系统的快速分析，以区分肿瘤边缘、健康组织、边界和恶性组织，将大大提高肿瘤手术的质量。组织切除量将得到优化，术中冷冻切片和时间延迟可以最小化，并且可以实现更好的闭合和重建。

创伤外科

现状

机器学习算法预测了踝关节骨折固定手术中开放复位内固定（ORIF）的术后短期结果 [30]。

潜在的

上颌骨外伤中隐匿性骨折的发生率很高，这是外科医生关注的问题。漏诊骨折和相关骨折在常规 X 线片上很难发现，经常困扰临床医生。机器学习系统可以有助于检测和隔离隐匿性骨折 [31]。

AI 深度学习指导从针对给定面部创伤的各种硬件选项中选择最合适的内固定，并通过分析患者射线照片、年龄和病史，根据医学文献进行个性化，例如，使用的钢板类型、螺钉数量、钢板与方头螺钉、一点固定与二点/三点固定。这种预测智能将防止因固定不当或过度热情而导致的畸形愈合和不对称等并发症。

AI 计算机视觉在面部创伤（多发性面部骨折）和弹道（高能撕脱伤、爆炸诱发）面部损伤中尤其有用。使用计算机视觉导航手术系统，可以大大提高全面部创伤测序和隐匿性创伤诊断以及组织（软组织和硬组织）丢失程度的挑战。

全面的眼眶和眼部创伤人工智能工具将有助于同步眼科和颌面科团队之间的患者护理，从而实现快速护理。眼眶骨折通常发生在鼻筛骨和/或颧骨腋窝复合体的交界处。此外，在手术复位这些骨折之前，必须对视力、视野、眼底镜检查、乳头反应和眼球运动进行彻底的眼科评估。人工智能工具可以被训练以准确地进行光学测试，并对新手外科医生或创伤团队有很大的帮助，因为他们可以有限地使用眼科设备。

杂症

面神经麻痹分级由于创伤、感染、医源性原因和腮腺/颞下颌关节区域手术并发症导致的面神经麻痹，导致单侧面神经麻痹/贝尔麻痹样状态。此类病例的处理取决于神经元损伤程度的有效分级。目前没有客观可靠的面神经麻痹分类。使用患者执行指定面部动作以测试神经功能的视频片段可以作为人工神经网络的数据，以有效确定面部神经麻痹的程度/等级 [34] 。

疼痛控制颌面外科特别是牙科和医学领域的桥梁专业，其程序通常包括局部麻醉和全身麻醉以及有意识镇静和动态麻醉。据报道，人工智能通过基于体内药物药代动力学的机器学习预测自动药物输送来实现最佳麻醉输送。神经网络已用于预测麻醉恢复率。正在测试机器学习技术，以监测意识镇静和警惕呼吸抑制水平，并自动对患者术前 ASA（敏锐度）状态进行分类 [35] 。

未来清单

颌面外科医生

外科数据科学需要大量的注释数据来训练人工智能系统。这导致了一个独特的外科数据科学领域的出现。这涉及到使用人工智能工具来捕捉、组织和分析，进而为人工智能系统提供燃料，以提供人工智能增强的手术室体验。此类人工智能系统将通过提供实时决策支持、上下文感知辅助和认知机器人技术来提高外科护理质量 [36] 。

手术场景分析

最近的研究集中于手术工具的 AI 分析，以实现自动化手术（机器人）器械的使用。然而，未来该方法将扩展到整个手术场景分析。这可以转化为任何东西，从虚拟的第二助手到实时的术中预测设备 [37] 。

外科控制塔（SCT）

人工智能是通过研究人体检测姿势估计和通过壁挂/天花板摄像机图像数据进行手术过程监控而开发的。SCT 将作为对手术团队进行全面协助的中央控制。可以分析正在进行的 OR 状态，以提供智能决策和任务自动化，并减少复杂情况下的错误。SCT 还可以作为训练和评估数据的视频库 [38] 。

结论

从公元前 2700 年埃及人 Edwin Smith Papyrus [39] 对下颌骨骨折治疗的首次报道到最

先进的“Fibula Jaw in a Day”颌面重建 [40]，颌面外科的发展受到了时间的影响。幸运的是，从医学中脱颖而出并与牙科合并成为一个独特的桥梁专业的兄弟会一直在不断升级和适应，以加强实践。培训模式正在从传统模式转变，最主要的影响是技术。人工智能在颌面外科的应用范围将见证该领域的各个方面实际应用——从人工智能驱动的拔牙自动化和机器人人工智能增强的手术器械，到机器学习支持的决策支持，在诊断疑难杂症和具有挑战性的医疗条件（如头颈遗传综合征）方面。应用是无限的，本章深入研究了当前的文献和 AI 集成颌面外科未来的潜力。

P905-P916;

牙科中的 AIM

Mauricio do Nascimento Gerhardt, Sohaib Shuaat, and Reinhilde Jacobs

摘要

数字牙科的急剧崛起和技术进步为人工智能 (AI) 的发展打开了大门。在过去的几年里，基于人工智能的牙科应用一直在不断发展，越来越多的研究所强调，现在它正在慢慢进入临床领域。作为医疗保健专业人员，牙医需要诊断、计划和做出临床决策，以便为患者提供充分的治疗和护理。所有这些阶段都是耗时的，依赖于观察者，并且容易出现人为错误。目前，在许多牙科专业中应用人工智能的研究已经验证了其在诊断和临床决策方面的应用。因此，人工智能的目标是将专业知识与计算机辅助系统相结合，以自动化复杂任务，模仿人类认知技能，并从数字数据中检索信息。牙科 AI 应用适用于所有牙科专业，包括牙颌面放射学、修复牙科、口腔颌面外科、正畸学、牙周病学、修复学、牙髓学和法医牙科。

尽管大多数研究和开发仍处于早期阶段，但目前在牙科领域的结果对未来的临床应用来说是令人鼓舞的。

本章概述了人工智能在牙科及其专业领域的应用现状。

关键词

数字化牙科、术前计划、锥束计算机断层扫描、口内扫描仪、全景放射线照相、放射学诊断、牙齿、下颌、面部

导论

牙医的日常任务包括收集患者信息、诊断以及计划和执行临床治疗程序。在第一次访问牙科诊所时，牙医需要详细收集和注释许多信息。之后，临床医生启动诊断阶段，结合临床和成像数据，将所有信息排列在牙科图表中。所有这些任务都很耗时，完全取决于专业人员对可能的小细节的关注。此外，充分整合这些多种信息和数据来源至关重要。只有在仔细收集和整合所有信息和诊断数据后，牙医才能开始进一步的临床治疗程序。

在过去的十年中，技术进步已经提供了一些解决方案，例如电子医疗和牙科记录，计划口腔康复的软件应用程序以及可视化计划治疗结果的增强现实。然而，这些解决方案仍然依赖于人为干预，需要大量时间。在这种情况下，重要的是开发更有效的软件，以允许可靠和自动化的行动，以帮助牙医进行评估、治疗计划和决策策略。这就是人工智能 (AI) 可以介入以促进数字化和自动化的地方。

人工智能在牙科领域正在迅速发展，过去几年中，全球研究和应用的数量急剧增加。牙科中的数据和信息的存储一直是日常临床程序的一部分，因为牙医经常收集成像数据，并伴有患者记录中的一般和牙科病史、全身状况和药物。如今，使用数字数据存储变得更加容易。这使得牙科成为实施人工智能及其子类、机器学习（ML）和深度学习（DL）等新技术的合适领域。技术的帮助可能在评估、诊断、计划、临床决策和结果预测方面发挥重要作用，从而改善预后。一些研究表明，基于人工智能的系统甚至能够超越牙科专家的表现[12]。

在牙科中，面部和牙齿美学是一个主要问题。随着牙科材料和技术的改进，患者对治疗结果的期望急剧增加。因此，在处理对人们的外观有影响的组织和解剖结构时，准确的诊断和治疗计划至关重要，并且考虑到患者的独特特征，还可以提供个性化的医疗保健。

因此，我们必须理解人工智能是一门进步的科学，它并不打算取代牙科专家，而是作为一个强大的盟友，通过牙医的同理心和护理，对病例进行更好、更快、更准确的评估和诊断。最后，它打算以更低的成本为患者提供更好的结果。这种技术和牙科医生的结合被称为增强智能[2, 3]。

本章的目的是让读者在牙科治疗的评估、诊断、计划和预测方面，在牙科治疗的评估、诊断、计划和预测方面，概述牙科人工智能在牙科中的应用现状（图 1）。

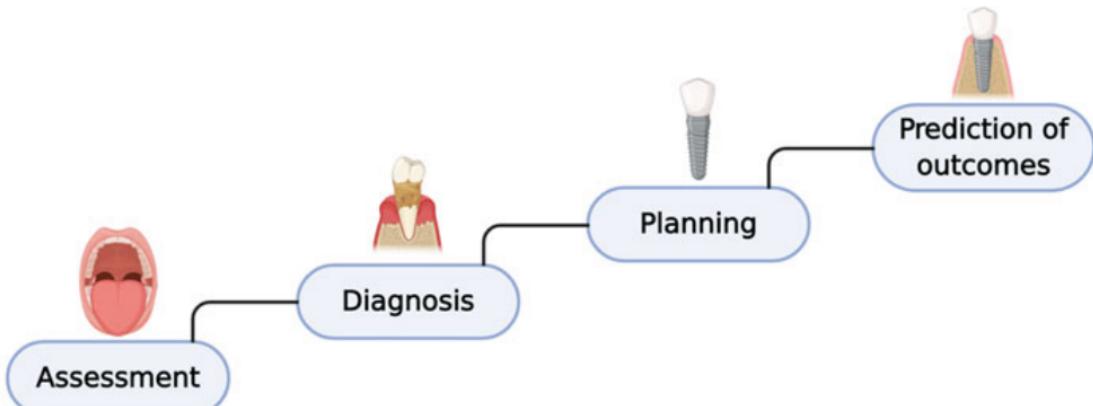


图 1 人工智能可以帮助牙科专家的牙科治疗步骤

牙科评估的人工智能

目前，口腔放射学是使用旨在改善牙科任务的计算机系统而受到最多关注的领域之一。由于计算机系统能够提取和量化人眼无法检测到的特征，从而将放射学变成一门客观的科学。在这种情况下，人工智能及其子类一直是为放射科医生分析不同类型图像中包含的信息量的绝佳工具。DL 已被应用于许多目的，例如优化低辐射剂量和较少散射或伪影的图像，分类，图像配准，分割，病变检测，图像检索和图像引导治疗。它还可用于纠正全景 X 线片上的技术定位误差和开发患者特异性成像方案[3]。

基于成像的数据也是进一步开发基于人工智能的应用程序的理想选择，主要用于治疗、手术计划和随访。训练自动化工具的第一步是解剖结构的分割。这通常是一项手动、耗时且依赖于操作员的任务，并且会受到人为错误的影响。这项工作的自动化可以改善诊断和治疗计划，虚拟手术计划，放射治疗计划和影像组学分析[4, 5]。

人工智能已被用于使用不同的放射成像方式识别牙齿。多项研究报告了使用自动化系统在根尖周、咬翼、全景和锥形束计算机断层扫描（CBCT）上有效识别和标记牙齿，其结果与专家获得的结果相当[69]。张等[6]和陈等[7]应用深度学习算法识别和标记口内 X 光片上的牙齿位置，并实现了高水平的精度。前者有 95.8%，后者的精度超过 90%。图佐夫等[8]还将 DL 应用于全景 X 光片上的牙齿检测和编号，灵敏度为 0.99，精度为 0.99 表示检测，灵

敏度为 0.98，特异性为 0.99 用于编号。莱特等[9]使用人工智能驱动的工具对全景 X 光片进行牙齿检测和分割，系统达到 98.9% 灵敏度和牙齿检测精度为 99.6%。他们还强调，该方法明显更快，比手动处理所需的时间少 67%。由于提供了一般概述，通常需要全景 X 线片进行病例评估。图 2 给出了用于牙齿检测和分割的 AI 驱动工具的示例。

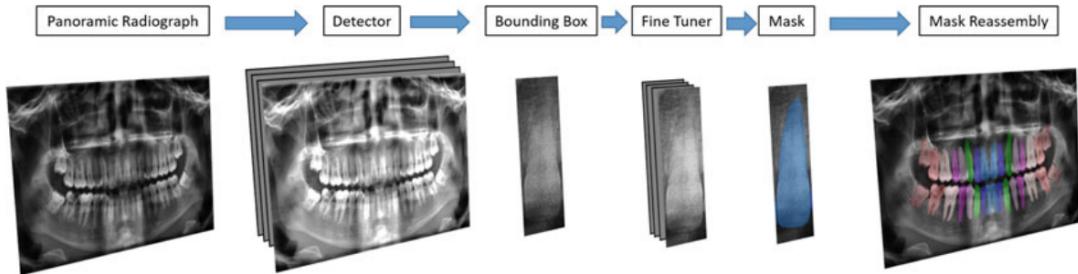


图 2 用于牙齿检测和分割的 AI 驱动工具的工作流程[4]

同样，使用 CT 和 CBCT 设备也对 AI 在牙齿识别和分割中的应用进行了三维探索[10-12]。图 3 展示了基于 AI 的牙颌面结构分割示例。所有这些结果表明，人工智能可以有效地应用于以节省时间的方式识别和标记牙科图表上的牙齿。

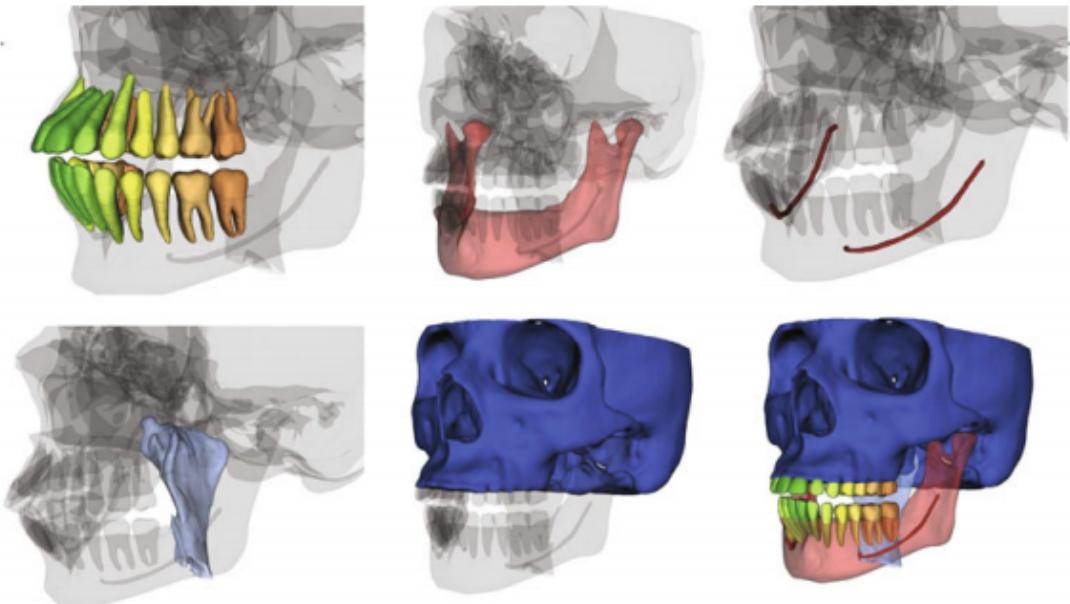


图 3 在 CBCT 图像上自动识别和标记的牙颌面结构 (Relu BV, 比利时鲁汶)

在种植牙领域，基于人工智能的种植牙识别可能是另一项重要任务，因为它们已成为口腔康复的首选治疗方法已有 30 多年的历史。从那时起，开发了许多不同的品牌和系统。随着全球植入物数量的增加，并发症的报告也有所增加。为了解决牙科夹具的问题，有必要识别种植体制造商及其修复连接系统。在某些情况下，种植体的识别很困难，因为牙科记录中可能缺少信息，或者患者正在就诊另一位不了解种植体特征的专业人员。因此，AI 可能是从放射影像检查中识别种植牙的有用工具[13-15]。高桥等[13]。在一项试点研究中应用了深度学习，以识别三家公司制造的六种不同系统中的植入物。结果表明，可以从全景 X 光片中识别种植牙。Hadj Said 等人[14]还应用 DL 识别来自三家制造商的六种类型的植入物，并实现了近 94% 的诊断准确率。Lee 和 Jeong[15]应用 DL 从全景和根尖周 X 光片中识别出三种相似的锥形种植牙，并将结果与牙周病医生的表现进行比较。

结果表明，DL 的曲线下面积 (AUC) 为 0.97 (95%置信区: 0.96, 0.98)，而专业值为 0.92 (95%置信区间: 0.91, 0.93)。前者的敏感性和特异性分别为 95% 和 97%，后者的敏感性和特异性分别为 88% 和 87%。尽管这两项研究的结果令人鼓舞，但它们要么依赖于一个小数据集，要么需要数据增强程序。因此，如果我们需要在常规临床实践中应用这些类型的人工智能工具，那么它们的性能应该在所有使用多种 2D 和 3D 成像模式和扫描参数的种植体系统中得到验证。

关于下颌种植体的放置，外科医生必须了解下颌管 (MC) 定位，以便在不损伤神经的情况下确定种植体尺寸。MC 的识别和分割通常由放射科医生在 3D CBCT 图像上手动进行。贾斯卡里等[4]。证明，与手动描述相比，使用 DL 分割 MC 显著减少了放射科医生在任务上花费的时间。准确度方面的结果在 0.5 毫米以内，用于约 90% 的 MC。展望未来，这种人工智能驱动的神经血管管检测和分割需要用更多的解剖变异性来自不同 CBCT 单位的图像进行训练。

在常规牙科实践中遇到的最常见情况之一是第三磨牙阻生，牙医需要预测牙齿是否会在口腔中正常萌出或需要拔牙。预测这些牙齿潜在萌出的两个重要因素是它们的角度和第二磨牙远端与支前缘之间的空间。放射科医生可以在全景 X 光片上进行这些测量。但是，与前面提到的所有其他任务一样，这是一项乏味且耗时的任务。Vranckx 等[16]。应用基于 AI 的算法在全景 X 光片上自动分割和确定下白齿的角度。他们测量了人工智能工具的性能以及与专家相比完成任务所需的时间。他们的结果表明，AI 工具在分割和角度计算方面具有良好的性能（交集，0.88）。此外，发现它的速度是手动测量的两倍。图 4 显示了全景 X 线片上下磨牙的自动分割和角度测量的示例。

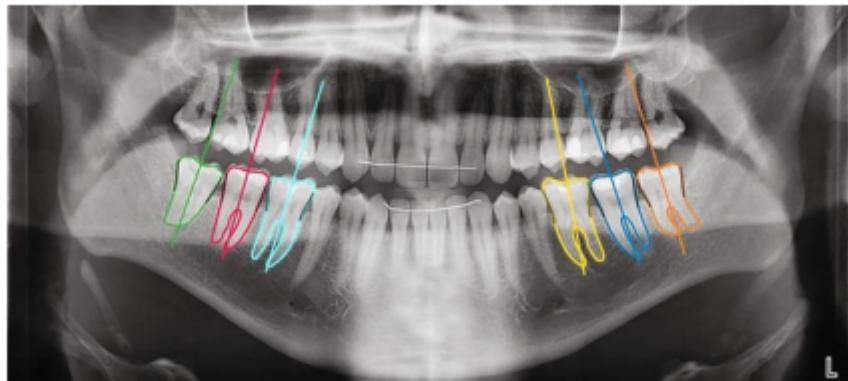


图 4 全景 X 线片上的自动牙齿角度测量[4]

在正畸治疗方面，通常会进行侧位头影测量分析，包括识别骨骼和软组织标志，目的是评估各种角度、距离和比率，并计划和预测治疗结果。追踪这些地标是一项繁琐的活动，手动识别仍然是黄金标准。这项任务取决于观察者，并且根据正畸医生的经验水平而有所不同。对标志的不准确识别可能导致正畸治疗的决策不当。已经证明，AI 和 DL 算法具有识别解剖标志物的潜力，并可准确检测头影 X 线片上的骨骼分类[17-20]。在评估正畸病例时，确定实际年龄和估计患者的骨骼成熟度也很重要。这两个因素对于评估可能影响正畸计划的剩余生长和发展潜力都很重要。一些研究调查了使用 AI 来执行这些任务[21]。上述所有任务，如射线照相分析和识别解剖标志，对于临床医生和放射科医生来说都是耗时的任务。因此，这些任务的自动化将有助于临床医生，因为在填写牙科图表或详细阐述放射成像报告时不需要人工干预。

在牙周病学中，牙周炎是口腔中最常见的疾病之一，可导致牙槽骨质流失和无牙颌。这反过来会影响患者的生活质量并增加医疗成本，从而需要早期和准确的诊断。有证据表明，

只有少数研究应用人工智能来诊断牙周疾病。Krois 等人[22]和 Kim 等人[23]提出了一种基于卷积神经网络 (CNN) 的网络，用于检测全景 X 光片上的牙周骨质流失。两项研究均显示出不同的发现，其中 Krois 等人[22]。与检查员 (0.76) 相比，CNN (0.67) 的准确度更高，没有任何显著差异，而 Kim 等人[23]建议 CNN 的准确度更高 (0.75)，牙科临床医生的平均得分为 0.69。在另一项研究中，Lee 等人 [24] 开发了一种 CNN 算法，用于诊断和预测根尖周围放射成像图像上的骨质流失。他们发现该网络与经验丰富的牙周病医生一样有效。最近，在另一项研究中，Chang [25]等人应用 CNN 进行诊断，并根据全景 X 线片分期牙周炎。他们证明了与没有经验的放射科医生相比，CNN 和经验丰富的放射科医生之间的高精度和更强的相关性，从而证实了其有效性和可靠性。由于 2D 射线照片提供不同程度的放大倍率和失真，因此，在日常实践中使用这些系统之前，需要进一步优化图像以训练 CNN 算法。

在牙髓病学中，顶端孔的定位对于准确确定牙髓管的工作长度 (WL) 至关重要。因此，不准确的 WL 可能导致牙髓去除不完全或根尖周组织侵入。萨吉里等[26]。进行体外研究并应用人工神经网络 (ANN) 评估文件尖端相对于小顶端孔的位置并确定 X 光片上的 WL。他们将 AI 性能与专业评估进行了比较，发现基于 AI 的网络在 96% 的病例中正确确定了 WL，而牙髓科医生则为 76%。基于这些发现，人工智能被证明是定位顶端孔和确定 WL 的绝佳辅助工具。

人工智能可以发挥重要作用的一个相对较新的领域是法医牙科。有时，只有通过分析牙齿遗骸才能识别飞机失事等重大灾难的受害者，法医必须花费大量时间来识别和分析数据。同样，法医牙科的另一个常见任务是通过评估第三磨牙发育阶段来估计青少年或年轻成人的年龄；然而，这种评估容易受到人为错误和观察者变异的影响。牙齿遗骸的自动识别、第三磨牙阶段的估计以及基于全景 X 线片的形态测量参数识别可能有助于区分一个人的身份、年龄和性别。在上述领域已经进行了一些基于人工智能的早期研究，从而证实在不久的将来，人工智能可能被认为在解决灾害和年龄估计方面有用[27, 28]。

牙科诊断的人工智能

AI 在牙科诊断中观察到的最有益影响是改善了检测可能被放射科医生或临床医生误诊或忽略的病变，例如龋齿、根尖病变、干燥综合征、鼻窦炎、颈淋巴结转移的结外延伸和骨质疏松症[3]。

龋齿是一种常见的慢性传染病，影响着世界各地的许多人。尽管在过去几年中，保存牙齿的修复方法有了显着改善，但诊断方法却没有发生同样的情况。如果在早期发现牙洞，则消退预后更乐观。然而，在存在深裂、近端紧密接触和修复体中的继发性病变的情况下，检测可能很困难。尽管牙科射线照相和使用牙科探头可以被认为是检测龋齿的绝佳工具，但最终诊断往往基于经验证据。基于人工智能的方法已被应用于检测龋齿，这可能代表了向诊断此类病变的客观方法的重要转变。德维托等[29]。应用 AI 模型在咬翼 X 光片上检测近端龋齿。结果表明，人工智能能够将诊断准确性提高 39。与审查员的平均值相比为 4%。Lee 等 [30]。还应用了基于深度学习的 CNN 来检测根尖周围 X 线片上的龋齿，该算法的平均 AUC 为 0.89。这些结果表明，使用人工智能对于检测牙医看到的最常见的病变之一是多么有益。此外，允许将完全取决于专业人员经验的主观诊断观点转变为客观观点。

裂缝或牙根骨折的正确诊断对于制定准确的治疗策略至关重要，这可能涉及拔牙或更保守的方法。即使在肉眼可能看不到骨折的 CBCT 图像上，这项任务也很困难，导致误诊和治疗方法不准确。一些研究调查了人工智能在有效检测根部骨折方面的应用[31, 32]。乔哈里等[31]。应用概率神经网络，使用根尖周和 CBCTX 光片检测完整和牙髓治疗牙齿上的垂直根部骨折 (VRF)。结果显示准确度高达 96.6%，作者得出结论，设计的神经网络可用于诊断

完整和牙髓治疗牙齿上的 VRF。在评估 VRF 时，CBCT 图像比根尖周围 X 线片更有效。Fukuda 等[32]。还应用 CNN 检测全景 X 光片上的 VRF。AI 模型的精度为 0.93，并且发现它比专家更有效。

如果不及早发现或治疗，一些口腔疾病可能会危及生命。在这些疾病中，口腔颌面癌位于金字塔的顶端，死亡率很高。口腔恶性肿瘤是口腔鳞状细胞癌（OSCC），在世界范围内发病率很高，大多数病例在晚期被诊断出来，从而降低了患者的生存率。奥伯维尔等[33]。评估了一种通过将 DL 方法应用于共聚焦激光显微镜内窥镜（CLE）成像来诊断口腔破裂的自动方法。他们将这种方法与基于纹理特征的 ML 方法进行了比较，并获得了 0.96 的 AUC，平均精度 88.3%。作者强调，这种非侵入性方法具有高放大倍率和更好的深度穿透的优点，可以诊断表面以下 100um 的恶性组织。自动实时识别口腔粘膜中潜在的恶性部位可能有助于加快治疗速度并改善预后。

一些颌骨肿瘤与侵袭性和侵袭性有关；然而，它们表现出相似的放射学特征。因此，治疗是基于鉴别诊断，而不是可能对治疗计划产生负面影响的绝对诊断。例如，角化囊性牙源性肿瘤（KCOTs）不需要根治性颌骨分割，而是像其他囊性病变一样需要剜除。另一方面，釉质母细胞瘤需要广泛的手术切除和重建。这两种肿瘤之间的适当诊断和放射学鉴别可以指导口腔颌面外科医生精确计划治疗。Poedjiastoeti 和 Suebnukarn[34]应用 CNN 对数字全景 X 光片上的 KCOTs 和釉母细胞瘤进行分类。结果显示准确率为 83%，与专家（82.9%）相当，但速度要快得多（38 秒对 23.1 分钟）。

用于治疗计划的人工智能

人工智能对于牙科规划也很有用。对于正畸治疗，在安装固定矫治器之前，可应用 AI 来决定是否需要正畸治疗或何时拔牙[3537]。此外，正畸治疗可能需要手术干预，例如正颌手术，以纠正上颌骨和下颌骨的骨骼异常，并且必须仔细和果断地计划。已经证明，AI 可以根据侧位头影在手术和非手术病例之间做出决定[38]。

在手术治疗的情况下，技术的进步使成像数据与虚拟模型相结合成为可能。此过程需要通过识别公共参考点来准确叠加数据。尽管已经提出了一些内在和外在的注册方法，但是，它非常耗时，并且需要某些软件程序的特定知识才能完成这样的任务。同时，可以通过 AI 以用户友好和省时的方式实现配准，从而可以准确测量和制造用于正颌手术，口腔颌面手术器具，种植牙插入指南，导航手术和许多其他应用的数字夹板。

用于牙科结果预测的人工智能

最重要的预测活动之一肯定是在疾病转移之前的早期阶段检测疾病，例如口腔癌，为患者提供足够的治疗和良好的预后。DL 可以应用于此目的，因为它从训练数据中学习特征，以对看不见的数据进行预测。

患者经常向牙医询问治疗的结果，这是一个很难提供的答案，因为预测结果受到许多因素的影响，这些因素在给予患者适当的反应之前需要一定数量的数据和统计分析。然而，一些研究侧重于使用 AI 来预测治疗的结局，例如拔除第三磨牙后的术后肿胀[39]。

对于正畸病例，根据混合牙列阶段的情况，正畸医生需要为未萌出的牙齿保存或打开空间。[40]的一项研究调查了使用 ANN 来预测混合牙列儿童未萌出的犬齿和前臼齿的大小。ANN 选择下颌第一磨牙和门牙以及上颌中央门牙分别预测下颌骨和上颌骨中犬齿和前磨牙的宽度。预测误差率和最大高估/低估率小于线性回归[40]。

正颌手术旨在纠正常规正畸治疗无法修复的严重牙颌面差异。这样的程序改变了面部美

学,这是患者的主要关注点。基于人工智能的方法已被应用于预测接受正颌手术的患者的面部软组织变化。卢等[41]。应用 ANN 改善了正颌手术后患者个人资料视频图像的面部变化预测。ANN 将预测能力提高了 80%以上。对手术结果的准确预测可能会改善治疗计划,并为患者提供手术后变化的准确描述。

AI 已广泛应用于保守牙科、牙科种植学和修复学,用于预测修复体的寿命及其潜在的脱粘,预测陶瓷配方以匹配天然牙齿的颜色和种植体治疗成功率以及全口义齿康复后的软组织转化[42-46]。预测结果的能力基于许多因素,例如[1]所需修复体的特征, [2]患者的个人习惯, [3]牙医的能力以及[4]修复材料的特性。用人工智能预测所有这些因素可能会改善决策过程。前进的道路是在更广泛的层面上实施这些预测性人工智能驱动的网络,因为它能够通过减少牙医费力的日常任务、预测结果并允许基于基于 AI 的预测模型的准确治疗计划来简化牙科护理。表 1 提供了本章报告的人工智能应用的摘要。

表 1 人工智能在牙科中的应用摘要

研究	领域	应用
Zhang et al (2018) [6]	口腔颌面放射学	口腔根尖周片对牙齿的识别
Chen et al (2019) [7]	口腔颌面放射学	口内牙周 X 线片上牙齿的检测和编号
Tuz of fetal. (2019) [8]	口腔颌面放射学	全景 X 光片上牙齿的自动检测和编号
Leite et al (2020) [9]	口腔颌面放射学	全景照片上牙齿的自动检测与分割
Hos nta lab et al (2010) [10]	口腔颌面部	多层螺旋 CT 对牙齿的分类和编号
Miki et al (2017) [11]	放射学口腔颌面放射学	CBC T 图像上牙齿的图像分类
Lahoud et al (2021) [12]	口腔颌面放射学	基于 CBC-T 图像的牙齿三维分割
Takahashi et al (2020) [13]	种植学	牙种植体在全景 X 线片上的识别
Had j Said et al (2020) [14]	种植学	牙种植体在 X 线片上的识别
Lee and Jeong (2020) [15]	种植学	牙种植体在根尖周片上的识别
V ranck x et al (2020) [16]	口腔颌面放射学	下颌磨牙角度的测量
Park et al (2019) [17]	口腔颌面放射学	头影测量地标的识别
Kunze tal. (2020) [18]	口腔颌面放射学	头影测量地标的识别
Hwang et al (2020) [19]	口腔颌面放射学	头影测量地标的识别
Yuet al. (2020) [20]	口腔颌面放射学	头影测量 X 线片上的骨骼自动分类
Kok et al (2019) [21]	正畸	生长期的自动判定
Kro is et al (2019) [22]	牙周病学	牙周骨丢失的牙体全景图检测
Ki metal. (2019) [23]	牙周病学	牙周骨丢失的牙体全景图检测
Lee et al (2018) [24]	牙周病学	根尖周片对骨丢失的诊断和预测
Change tal. (2020) [25]	牙周病学	牙周骨丢失的 X 线片诊断及分期
Saghiri et al (2012) [26]	牙髓学	小孔的定位
De To be letal. (2017) [27]	法医牙科	下颌第三磨牙期的自动判定

Patil et al (2020) [28]	法医牙科	基于下颌骨形态参数的性别判定
Devito et al (2008) [29]	汽车学	邻面龋病的诊断
Lee et al (2018) [30]	汽车学	龋病的检测与诊断
Johari et al (2017) [31]	修复性牙科	牙根纵裂的检测
Fukuda et al (2020) [32]	修复性牙科	牙根纵裂的全景 X 线片检测
Au breville et al (2017) [33]	电子口腔学	口腔癌组织阳离子的自动分类
Po edjia stoeti and Suebnukarn (2018) [34]	电子口腔学	颌骨肿瘤的诊断
Than at horn wong (2018) [35]	正畸	评估正畸治疗的必要性
Xie et al (2010) [36]	正畸	正畸治疗前拔牙必要性的评估
Jung and Kim (2016) [37]	正畸	正畸治疗前拔牙必要性的评估
Choi et al (2019) [38]	正畸	诊断正颌外科手术的必要性
Zhang et al (2018) [39]	口腔颌面部	下颌第三磨牙拔除术后肿胀的预测
Mog him i et al (2012) [40]	正畸	未萌出牙齿大小的预测
Chien-Hsu net al. (2009) [41]	口腔颌面部	面部侧貌手术后变化的预测
Aliaga et al (2015) [42]	修复性牙科	口腔修复体寿命的预测
Yamaguchi et al (2019) [43]	修复性牙科	CAD/CAM 修复体脱粘的预测
Alar if i et al (2018) [44]	种植学	种植体成功率的预测
Cheng et al (2015) [45]	口腔修复学	全口义齿修复后面部变形的预测
Wei et al (2018) [46]	口腔修复学	牙科配色陶瓷配方的预测

表 2 开发牙科 AI 解决方案的公司

公司	国家	描述
BubblesEndo	美国	改变牙髓诊断的人工智能牙科初创公司。名为“Endo genie”的产品是一种基于人工智能的牙髓应用程序，可帮助临床医生在检查过程中组织他们的发现，分析牙科 X 光片并突出显示其发现。该应用程序还将成像分析与临床试验的结果相结合，以提供最可能的临床场景网站： https://www.endogenie.ai
CellmatiQ	德国	该公司的使命是将复杂的图像分析扩展到人类能力之外，自动检测和分类结构模式，选择基于图像的工作流程的处理时间和数量，并以客观和有效的方法改进视觉评估。在牙科解决方案中，可用的模块包括头影测量分析模块和图像通道类别 IF 阳离子。目前，该公司正在开发用于牙科全景 X 线片的病理指示、龋齿检测和颈部 vertebral maturation Website: https://cellmatiq.com 年龄确定的解决方案

Dental Intelligence	美国	该公司的目标是帮助牙科医生提供更多更好的护理，并改善团队的协作和绩效，提高牙科医生的整体健康和效率。他们的牙科解决方案是一个连接到诊所管理软件并跟踪活动并分析、发现和发送信号的模块，以确保拯救我，帮助更多的患者，并提高产量。网站： https://www.dentlintel.com
Dental Monitoring	法国	这家法国公司希望重塑患者体验，彻底改变口腔保健的提供方式，将牙科实践扩展到虚拟环境。他们的解决方案包括监控患者治疗的应用程序，生成超现实微笑预测的人工智能技术，以及宠物任务的自动化。网站： http://denta-monitor oring.com
dental Xr.al	德国	该公司将机器学习、软件工程和专家研究结合在一起，致力于支持牙医为他们的患者做出尽可能好的诊断和治疗决策。他们的技术支持牙医在牙科放射照片中检测病理和修复。网站： https://dentalxr.ai
Denti AI	加拿大	基于云的人工智能应用于解释牙齿 1 图像，并被开发为工作的“第二意见”，帮助识别未诊断的情况和快速制定牙科图表。网站： https://www.delicai
Diagno cat	俄罗斯	一个允许上传、存储和共享 Denali 图像的平台，该平台使用 BULT-INAI 技术自动检测常见的牙科疾病和病理。网站： https://diagno.cat
glidewell.io	美国	一家承诺改善患者护理并增加利润的公司，提供的解决方案使用机器学习技术和 AI 算法来创建定制的牙冠方案，并确保使用颜色编码进行准确扫描以显示主题区域。网站： glidewell.io
Kapanu AG	瑞士	这家瑞士公司旨在帮助牙科社区实现高质量和多 DS 的学科服务和产品。其中，有软件应用程序 VSU 利用增强现实技术实现美容牙科改造的可能性，并预览潜在的正畸治疗结果。网站： https://kapanu.com
MMG Fusion	美国	MMG Fusion 解决方案连接到牙科实践管理软件，帮助寻找可用于归档椅子的时间，增加冰块生产和收入，减少员工管理费用，并创造出色的患者体验。网站： http://mmgfusion.com

NovoDynamics	美国	<p>Novo Dynamic 产品将大型、完全不同的数据集转换为可操作的信息和洞察力，帮助组织使用 AI 解决方案做出更好的决策，以应对高度复杂的挑战。在他们的产品中，有用于射线分析的解决方案，以及针对提交给牙科 insurances Website: https://www.novohealthdental.com 的异常索赔的增强和检测</p>
Orca Dental AI	以色列	<p>这家公司的使命是为牙科医生提供取得突破性住院护理所需的工具。该公司将临床经验和人工智能技术相结合，创建诊断报告、治疗方案建议和智能临床预测。在他们的产品中，有自动头影测量分析、呼吸道分析、CBC TS 中的下颌神经管识别和龋齿检测网站： https://www.orca-ai.com</p>
Overjet	美国	<p>这家在哈佛创新实验室孵化的公司致力于解决牙科领域最具挑战性的问题。他们拥有保险索赔审查、管理任务自动化以及牙齿解剖、口腔疾病和修复质量识别的模块。网站： https://www.overjet.ai</p>
Pearl	美国	<p>基于人工智能可以成为牙科助手的概念，珀尔的目标是迎来新一波人工智能支持的工具，例如对每一张 X 光片的第二意见，牙科实践的管理，间接修复边缘的划定，以及自动保险索赔审查。网站： https://hellopearl.com</p>
Relu BV	比利时	<p>Re Luisa Young 是一家联合人工智能工程和研究专业知识开发临床应用的公司。在他们的解决方案中，有一些模块用于在 CBC T 扫描上自动分割牙齿，在牙科全景图上预测第三磨牙萌出时的磨牙角度计算，下牙槽神经的检测，以及咽部、下颌和头骨的分割。网站：https://relu</p>
Simplifeye	美国	<p>该公司专注于提供尖端软件技术和世界一流的服务。他们提供在线解决方案，随时随地接触到患者，并访问他们在云网站上可用的信息。网站： https://simplifeye.co</p>
VideaHealth	美国	<p>这家总部位于波士顿的公司为图像分析、诊断、治疗计划和保险公司索赔处理提供人工智能驱动的工具。网站：https://www.videa.ai</p>

结束语

牙科许多专业的自动化工具的开发仍处于早期阶段;然而,结果是有希望的。表 1 总结了一些 AI 应用。大多数研究表明,人工智能有能力在更短的时间内完成与专家相同或更好的人类活动,而不必担心观察者的变化。

目前,我们看到人工智能作为调制解调器牙科助理进入牙科领域。许多公司一直致力于开发基于人工智能的牙科解决方案(表 2)。未来,使用大型数据集训练用于各种临床情况的系统将能够从输入中识别各种特征,例如口内和口外 X 光片,照片,断层扫描和数字模型。这将允许自动评估患者的口腔状况,诊断并建议最佳预后的最佳治疗方案。此外,它将为根据每个案例的特定特征预测结果铺平道路。

P919-P933;

人工智能在胃肠病学中的应用

Inga Strumke, Steven A. Hicks, Vajira Thambawita, Debesh Jha, Sravanthi Parasa, Michael A. Riegler, and Pal Halvorsen

摘要

长期以来，内窥镜检查的圣杯一直是使用人工智能（AI）进行辅助诊断的。计算机硬件的最新发展现在使技术能够为临床医生提供计算机辅助诊断（CAD）系统的有前途的工具。然而，创建可行的模型或架构，训练它们，并评估它们在人类水平上的诊断能力，都是复杂的任务。这是目前一个活跃的研究领域，已经提出了许多有前途的方法。在本章中，我们将概述该主题。这包括对当前医疗挑战的描述，然后是该领域最常用的方法的描述。我们还介绍了针对其中一些挑战的研究结果，并就未决问题和正在进行的工作进行了讨论。希望这将启发并使读者能够在未来开发胃肠病学的 CAD 系统。

关键词

胃肠内窥镜检查 人工智能神经网络 手工制作功能 异常检测 语义分割 性能

导论

在人类胃肠道（GI）中可以找到许多异常的粘膜发现，从轻微的烦恼到高度致命的疾病。例如，根据国际癌症研究机构的数据，大约 3.全世界每年发现 500 万例管腔 GI 癌（食管癌、胃癌、结直肠癌）[41]。这些癌症对社会的健康构成重大挑战，死亡率约为 6365%，导致约 2. 每年有 200 万人死亡[19, 41]。总体而言，结直肠癌（CRC）是女性和男性癌症死亡的第三大常见原因[104]，其他最常见的胃肠道癌症是胃癌、肝癌、胰腺癌和食管癌[18]。

对于胃肠道疾病的诊断和治疗，胃肠道内镜检查是用于检查胃肠道异常的金标准程序，在一定程度上，可以通过改善内镜表现和在高发区域进行高质量的系统筛查来预防胃肠道疾病[19]。然而，尽管内镜在过去二十年中取得了重大的技术进步，但内镜检查的一个主要局限性是内镜操作人员的差异，这取决于手术技能、知觉因素、人格特征、经验、知识和态度缺陷[34]。这意味着观察者间黏膜病变的检测和评估存在很大差异[64, 108]。例如，这会导致结肠镜检查期间息肉漏诊率平均为 20%[52]。所有这些因素在某种程度上都可能通过大量的教育努力来缓解，但不能消除[88]。

在这种情况下，使用计算机进行辅助诊断长期以来一直是圣杯。计算机硬件的发展使计算要求高但有前途的技术成为可能，如 AI，更具体地说是其子领域机器学习（ML），为临床医生提供潜在的高度准确和高效的计算机辅助诊断（CAD）系统，为医疗保健专业人员提供大规模提供优质护理所需的工具[86, 102]。机器学习的核心是使用算法来解析数据，从中学习，然后进行预测，在医学领域，这意味着检测、分割、评估或分类疾病。但是，在创建和改进自动诊断算法方面，存在一些需要解决的问题。开发和评估计算机在人类水平上的诊断能力是一项复杂的任务，潜在的成功取决于各种因素，而不仅仅是确定算法的准确性。大约十年来，这些挑战一直是一个活跃的研究领域，并且已经发表了大量有希望的结果。

在本章中，我们描述了当前在实现有效的基于计算机的数字助理系统的过程中面临的挑战。特别是，我们专注于胃肠道内窥镜检查。我们提供了采用各种技术的拟议方法和工具的

示例，确定了当前的挑战，并为 CAD 系统的未来开发和评估提供了提示。

胃肠道内窥镜检查

为了检查食道、胃、十二指肠（上消化道）和大肠和直肠（下消化道），将一根长而灵活的管分别插入口腔和直肠。管子尖端的微型摄像机允许医生实时查看胃肠道内部，如图 1.a 和 1.b。

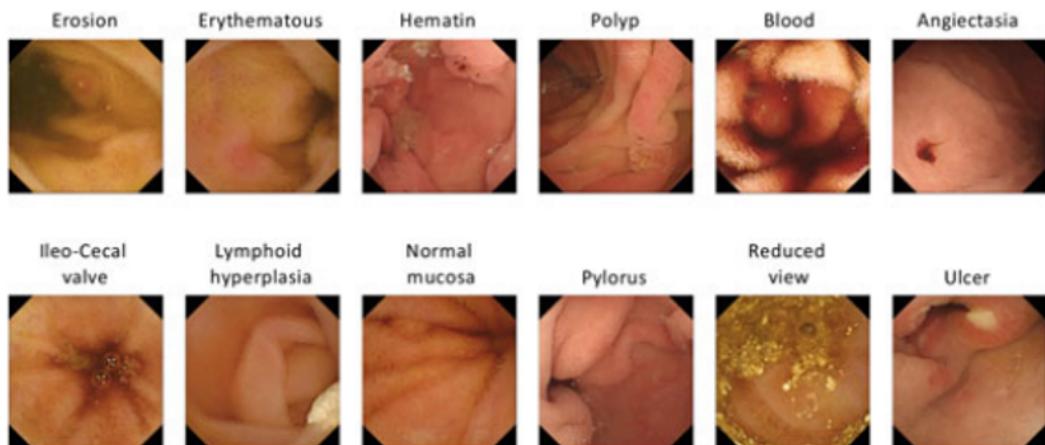
由于其解剖位置，这种柔性内窥镜不太容易检查小肠。为了更容易进入胃肠道的这些区域，已引入视频胶囊内窥镜检查（VCE）[22]作为替代检查方法[25]。VCE 由一个包含一个或多个广角相机的小胶囊组成。胶囊被患者吞咽，并在通过胃肠道时捕获视频。视频被提取出来，医学专家在手术后以可能乏味且耗时的过程对其进行评估，参见图 1.c。



(A) 上消化道、食道、胃(胃镜检查)



(B) 下消化道、大肠(结肠镜)



(C) 下消化道、小肠(胶囊内窥镜)

图 1 胃肠道中各种发现的示例，包括解剖标志，病理发现，正常粘膜，治疗干预和医疗器械
[14, 96]

尽管这些检查程序允许临床医生检测胃肠道异常，但仍有很大的改进空间。查看图 1 中描述的可能发现，很明显，很难检测和分类胃肠道各个部位可能发现的各种异常，无论是在胃镜检查或结肠镜检查期间，还是在 VCE 视频的后期分析中。此外，常规内镜检查[34, 52, 64, 108]和胶囊内镜检查[20, 88]均存在较大的操作者差异和异常漏诊率。

因此，希望自动分析可以帮助医学专家进行实时异常检测，消除变异并提高检测率。此外，分析数小时的 VCE 视频，与目前医务人员通常执行的 4560 分钟容易出错的快进分析相比，通过快速计算机在几分钟内分析 412 小时的长视频，在节省医学专家时间方面也有很大的潜力。从分析的角度来看，这种 CAD 系统有两个重要的要求：

1. 分析中的高检测或分割性能对于解决较大的人为失误率和变异性非常重要。它通常根据精度、灵敏度（召回率）、特异性、准确性、F1 分数、马修斯相关系数（MCC）或类似指标来衡量[98]。此要求旨在正确查找所有异常，即检测所有发现，没有误报或假阴。有关指标的更详细讨论，请参阅“指标和评估”部分。

2. 一个经常被忽视的要求是快速处理，以便在内窥镜检查期间提供实时反馈，或者在 VCE 的情况下，更高的分析规模和对相同数量的处理资源的更快反馈。

此外，为了在临床环境中部署，所有组件都需要集成到管道中，通过自动分析从内窥镜设备捕获视频或帧，以便为临床医生提供视觉反馈（并可能还可能帮助根据医疗标准生成检查报告）。该系统还必须易于集成到当前的检查程序中并可与当前的检查程序一起使用，当然，各种组件必须符合医疗隐私和安全法规。

现有方法

如上所述，已经提出了大量用于 GI 视频和图像自动分析的算法和模型。在这方面，当我们今天讨论胃肠道的 CAD 系统时，人们经常互换地谈论检测，定位和分割。因此，在这里，我们首先尝试区分以下术语：

检测是检测图像是否属于某种分类的操作。这可以是二进制“是或否”，用于图像或视频帧是否包含息肉的问题。它还包括将输入分类为多个类的系统。

定位是指向对象所在的图像，例如，使用某种类型的点标记或在感兴趣的对象周围制作边界框。

分割是进一步的一步，其中像素确定像素是否属于发现，例如，生成发现的精确分割掩码。

图 2 显示了检测、定位和检测的示例。由于本地化通常混合到检测和分割中，因此我们在这里重点介绍检测和分割。



图 2 指示发现的各种方法（左：检测 “仅” 显示图像;中心：全分割掩码以白色显示发现的所有像素部分;右：在发现周围形成矩形的边界框）

手工制作的基于特征的方法

早在人工智能和深度神经网络成功之前，GI 异常的自动检测就已经成为一个研究课题，使用现在通常称为传统计算机视觉和 ML 的方法，如 OpenCV [16] 和 LIRE [69] 等库。早在 1998 年，Krishnan 等人[59]就提出了在曲率分析中使用形状特征检测息肉。在随后的十年中，出现了各种使用形状，边缘，纹理和颜色特征混合的方法。例如，Alexandre 等人[2]使用支持向量机（SVM）在颜色图案上检测息肉。此外，使用 SVM，Ameling 等人[5]组合了纹理和颜色，Park 等人[75]在条件随机场分类器中使用了形状和纹理特征。

使用手工技术的两种最新方法是 PolypAlert[111]和 EIR [85]，作者还测量了分析时间，目的是能够在检查期间提供实时反馈。PolypAlert [111] 系统结合了边缘和纹理特征。息肉边缘检测算法主要依赖于从基于零件的多导数边缘截面轮廓中获得的边缘特征[110]。EIR [85] 系统将基于内容的相似性搜索与来自训练数据的统计分类器相结合。由于精度和灵敏度（召回率）以及算法速度之间的良好权衡，测试了大量图像特征[87]，最终结合了联合复合描述符特征和田村特征。然后使用基于搜索的分类器来确定图像是否包含某个类的查找结果。

包含早期示例方法的详细概述可以在 [85, 111] 中找到。然而，最近，深度学习方法已经超越了这些手工制作的方法，并完全取代了它们。

基于深度学习的方法

早在 2001 年，Karkanis 等人[53]就旨在使用人工神经网络架构支持的小波域上的纹理描述符检测内窥镜视频中的病变，尽管没有使用深层架构。这种早期方法在微小的数据集上进行了测试，在本例中为 8 张图像[53]。最近的方法通常基于深度学习架构，其中卷积神经网络（CNN）显然是最受欢迎的。

手工制作的特征依赖于提取图像的预定义属性，例如颜色、纹理或形状，而 CNN 是使用卷积和池化操作来自动学习哪些特征最相关的神经网络架构。CNN 在许多不同的任务上表现良好，如图像分类、图像中的对象检测和图像生成[56]。虽然它们主要用于图像分析，但它们也被证明在时间序列研究和视频分析中很有用。在医学领域，像 UNet[89]这样的架构在心脏病学，结肠镜检查和放射学等领域显示出有希望的结果[74, 119, 122]。这还包括胃肠病学，其中 CNN 目前是分析结肠镜视频的最先进技术。最常见的应用是息肉的检测和分割，许多基于 CNN 的方法显示出出色的结果[17, 50, 114]。这些方法也扩展到其他发现，如检测和分割溃疡[31]。此外，由于对医学图像和视频数据的访问有限，大多数方法都使用迁移学习。在迁移学习中，预训练模型被用作起点，并通过重新训练一些层和一些冻结层来针对给定数据集进行优化[82]。

用于胃肠道内窥镜图像分割的自动化 CAD 系统比提供“仅”异常检测更进一步。预测的分割掩码（见图 2）可以帮助指出需要进一步检查的图像（帧）中的感兴趣区域。然而，进行这样的每像素预测也是一项更复杂的任务。在这方面，到目前为止已经做了大量的工作，特别是针对息肉[32, 45, 47, 48, 50, 71, 81, 100, 109]、伪影[3]和内窥镜仪器[90]。一般来说，基于 CNN 的方法对较大的息肉表现良好。然而，该领域的主要挑战问题仍然与腺瘤性息肉或小而扁平的息肉有关。最近的研究针对较小的息肉[50, 63]；然而，这仍然是一个需要解决的开放挑战。

无监督和半监督方法

上面介绍的方法属于监督学习的范畴，这意味着我们在具有现有基本事实的数据集上训练模型。在本节中，我们将介绍新兴的无监督和半监督方法。

生成对抗网络（GAN）由 Goodfellow 等人[30]于 2014 年推出，在医学领域越来越受欢迎，用于生成合成数据。对原始 GAN 架构的不同改进，例如条件 GAN[72]，pix2pix [43]，CycleGAN [123]，StyleGANs [54, 55]，仅举几例，提出了不同的方法，从域转换到高清图像生成。医学领域的 ML 研究人员可以使用 GAN 模型生成合成数据，以应对与隐私、数据缺陷和数据注释相关的挑战。例如，Younghak 等人[93]使用条件 GAN 架构生成合成息肉图像，以提高深度学习系统检测结肠息肉的性能。这种方法仍处于早期阶段，尚未显示生成

的数据在多大程度上可以取代真实数据，并有助于提高性能和可共享性。

医学图像分析领域的另一种新兴方法是半监督学习。在这里，目标是从一小组标记数据与大量未标记数据相结合中学习。示例包括 [7, 67, 70, 116]。这些模型产生了有希望的结果，也有助于克服许多数据饥渴的方法所面临的标记数据不足的挑战。然而，这些方法在训练过程的早期阶段仍然面临着诸如低准确性和高熵等挑战。这些模型也正则化为高熵预测，因此很难实现高精度[117, 120]。看看这些挑战是否可以克服，以及结果在医学领域将证明有多大用处，这将是一件有趣的事情。

示例结果

高检测率或分割率对于临床相关性以及比较性能的典型方式非常重要。但是，由于数据集不同、设备不同等因素，无法直接比较纯数字。尽管如此，Tor 还是给出了一些最先进的性能迹象，我们给出了一组远非完整的示例，使用标准指标，如精度、灵敏度（召回率）、特异性、准确性、F1 分数和 MCC 进行检测；和骰子相似系数 (DSC)、并集相交 (IoU)、分割精度和灵敏度。现有方法的实质性概述可以在[61]中找到，其中包含 138 项不同的研究。表 I 中给出了不同指标的解释，并在“指标和评估”一节中进一步讨论。探索和比较不同方法的另一个来源是“竞赛和挑战”一节中讨论的流行的 GI 检测，分类和分割挑战。

表 2 给出了一些性能示例。查看数字，我们看到，在特定测试案例中，计算机应该处于得分高于 90% 的最佳专家的水平，即在胃肠道内窥镜检查期间可能是一个有用的数字助理。同样，表 3 提供了病变分割的示例结果，这些数字再次令人鼓舞，证明所使用的模型可用于医疗环境。然而，虽然取得的结果很有希望，但仍存在一些悬而未决的挑战，包括泛化性、过拟合、交叉数据集测试和结果的可解释性。此外，如表格所示，几乎没有任何现有研究报告系统的速度，这意味着很难评估系统在诊所提供实时分析的能力。

表 1 常用指标列表要定义每个指标，TP、FP、TN 和 FN 分别表示真阳性、假阳性、真阴性和假阴性

Formula	Description
$\text{accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$	Rate of correct classification. Ratio between correctly classified samples and all samples.
$\text{precision} = \frac{TP}{TP+FN}$	Proportion of retrieved samples which are relevant. Ratio between correctly classified positive samples and all samples classified as positive.
$\text{sensitivity} \text{ (also known as recall)} = \frac{TP}{TN+FN}$	Proportion of relevant samples which are retrieved. Ratio between correctly classified positive samples and all positive samples.
$\text{specificity} = \frac{TN}{TN+FP}$	Negative class sensitivity. Ratio between correctly classified negative samples and all negative samples.
$F1 = \frac{2 \times TP}{2 \times TP + FP + FN}$	Harmonic mean of the precision and sensitivity (recall).
$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)d(TP+FP+FN)}}$	Pearson's correlation coefficient [23] for binary classification.
$\text{IoU} \text{ (also known as Jaccard)} = \frac{TP}{TP+FP+FN}$	Similarity between sets from the size of the intersection divided by the size of the union.
$\text{DSC} = \frac{2TP}{2TP+FP+FN}$	Quotient of similarity of two sets. Semi-metric as it doesn't satisfy the triangle inequality. Related to the IoU via $\frac{S}{2-S}$.

表 2 不同方法的检测性能示例结果显示有希望的表现

Paper/system	Data set used	Sensitivity (recall)	Specificity	Accuracy	Precision	F1	MCC	Speed (fps)
Boughorbel [17]	MICCAI-challenge data sets	86.3	–	–	73.6	–	–	–
Kundu [60]	30 Own data set	95.2	98.3	97.9	88.4	–	–	–
Cho [21]	Seoul National University Hospital	>87	–	>93	–	–	–	–
Ghosh [29]	VCE videos data set	99.4	99.2	97.9	95.8	–	–	–
Bell [8]	CTC generating 4000 images per patients	89.8	75.5	–	–	–	–	–
Pogorelov [76]	Kvasir	83.9	98.5	97.2	84.1	85.6	82.8	46
Billah [13]	Colonoscopy & Endoscopy vision data set	98.7	98.2	98.3	–	–	–	–
Thambawita [99]	Kvasir	95.8	99.7	95.8	95.9	95.8	95.3	29

表 3 应用于不同数据集的不同分割方法的一些示例。我们可以清楚地看到，整体性能非常有希望（所有指标都在 70 到 95 之间）

Pape/system	Data set used	DSC	IoU (Jaccard)	Sensitivity (recall)	Precision	Speed (fps)
U-Net [89]	MICCAI-PhC-U373	–	92.0	–	–	–
PraNet [26]	CVC-ClinicDB	89.9	84.0	–	–	–
PolypSegNet [71]	CVC-ClinicDB	91.5	86.2	91.1	96.2	–
ResUNet++ [50]	CVC-ClinicDB	79.6	79.6	70.2	87.9	–
PraNet [26]	Kvasir-SEG	89.8	84.0	–	–	–
PolypSegNet [71]	Kvasir-SEG	88.7	82.5	84.5	91.7	–
ResUNet++ [50]	Kvasir-SEG	81.3	79.3	70.6	87.7	–
Double-UNet [45]	CVC-ClinicDB	92.4	86.1	84.6	96.0	–

开放问题和正在进行的研究

尽管许多已发表的论文都取得了令人印象深刻的结果，甚至超过了临床医生报告的平均检出率，但仍存在挑战和悬而未决的问题。

首先，例如，Thambawita 等人[98]提出了对特定数据集过度拟合和缺乏可推广性的问题。这意味着在一个数据集上表现良好的模型在另一个数据集上可能根本无法执行。此外，与其他深度神经网络一样，CNN 是黑匣子，很难理解为什么一个输入会产生特定的结果。也缺乏大型开放数据集，其中包含对不常见异常的注释，以及很少记录的发现来支持 CNN 等数据饥渴算法。在这里，我们详细阐述其中的一些未决问题。

有限的数据可用性

可用的医疗数据很少。然而，现代深度学习方法通常需要大量数据才能表现良好，而且通常，数据的变化越多，模型就越好，特别是对于监督学习模型。表 4 显示了胃肠道内窥镜检查领域可用的数据集。显然，与来自自然图像的数据集相比，用于训练和测试的图像数量很少。这是因为很难从医疗领域获取数据。由于法律限制和缺乏医务人员来手动提取和标记培训数据的繁琐过程，数据通常受到保护且不可用。这需要更好的数据共享流程和文化，或

者更有能力处理少量数据的方法。

这带来了几个基本挑战：数据量太小，无法训练鲁棒模型，并且由于过度拟合，呈现的结果可能看起来不错。此外，如果所有实验都对不同的数据进行，则很难比较结果，并且几乎不可能重现它们。因此，几乎不可能得出一种模型是否优于另一种模型的结论。因此，我们必须以更多和开放的数据集为目标。表 4 包含撰写本文时已知可用数据集的概述，为未来的实验提供了良好的起点。尽管如此，仍需要更多的数据，特别是包含病理结果的数据。

表 4 现有内窥镜数据集摘要。有关数据集的进一步讨论可在 [14, 96] 中找到。

Data set	Findings	Location	#Images	#Videos	Bounding box?	Segmentation mask?	Input size	VCE data?	Endoscopic device
Kvasir [77]	various	↑↓	8,000	—	—	—	variable	—	†
Nerthus [78]	stool (cleanliness)	↓	5,525	—	—	—	720 × 576	—	†
HyperKvasir [14]	various	↑↓	110,079	373	—	—	variable	—	†
KvasirInstrument [46]	instruments	↓	590	—	✓	✓	variable	—	†
Kvasir-SEG [49]	polyps	↓	1,000	—	✓	✓	variable	—	†
ASU-Mayo [97]	polyps	↓	18,781	—	—	—	variable	—	—
CVC-ClinicDB [10]	polyps	↓	612	—	—	—	384 × 288	—	‡*
CVC-ColonDB [11]	polyps	↓	380	—	—	—	574 × 500	—	—
ETIS Larib Polyp DB [95]	polyps	↓	196	—	—	—	1225 × 966	—	‡*
SUN Colonoscopy Video DB [73]	polyps	↓	158,690	—	✓	—	1080 × 1240	—	?
CVCVideoClinicDB [6, 12]	polyps	↓	11,954	—	—	—	384 × 288	—	—
CAD-CAP [65]	various		25,000	—	—	—	—	✓	—
KID [58]	various		2,371	47	—	—	—	✓	—
KvasirCapsule [96]	various	↓	4,820,739	118	✓	—	variable	✓	•

Location: ↑ = upper GI ↓ = lower GI

Device: † = ScopeGuide, Olympus ‡ = Olympus Q160ALandQ165L

* = Exera II videoprocessor • = Olympus EC-S10 endocapsule

一般化

该领域的一个开放问题是胃肠道内窥镜检查是 ML 模型的可推广性，即无论来源，设备等如何，它们都能在以前看不见的数据上表现良好。此类数据可以来自训练模型的同一分布，也可以来自不同的分布。新数据样本代表两者中的哪一个并不总是很清楚[101, 103]。尽管最近的一些研究解决了 ML 模型在息肉分类中的可推广性[45, 112]，但对于要部署到临床实践中的任何模型或系统，都必须解决这个问题。

评估模型在现实世界中是否可靠还需要跨数据集测试，以避免接受巧合地在一组特定数据集上运行良好的模型。模型开发人员通常不应访问最终测试数据，以避免在测试和开发过程中出现偏差。这个过程被称为数据盲法，是包括医学在内的许多研究领域的重要工具[80]。理想情况下，应根据与模型开发和测试期间使用的数据分开收集的数据测试模型的鲁棒性。

此外，应区分医学专家注释的数据（称为软地面真相）和基于医学测试标记的数据（称为硬地面真相），例如息肉的病理检查。软地面实况数据的质量受到医学注释员训练程度的限制，此类数据对于旨在自动化流程的训练模型最有用。另一方面，具有硬真实标签的数据也可用于自动化过程，具有避免注释器错误或模型偏差的额外好处，但它还可以用于获取新知识。请注意，如上所述，虽然注释每个图像非常耗时，但收集硬地面真实数据的要求更高，导致此类数据集的稀缺。

在目前的内窥镜检查实践中，不同的医院使用不同的内窥镜系统进行诊断和治疗。全球最常见的内窥镜系统是奥林巴斯（日本）、宾得 90i 系列（日本）、富士能（日本）和卡尔史托斯（德国）[57]。此外，不同的医疗机构有不同的协议。因此，设计可推广的 CAD 系统对于在各种机构中表现良好至关重要。此类系统应始终在多个数据集上进行测试。有关与跨数据集测试相关的挑战和优势的讨论可以在[98]中找到。

指标和评估

在创建临床使用模型时，评估性能是重要步骤，并且在很大程度上取决于指标的选择。如表 1 所示，常用的指标是精度、灵敏度（召回率）、特异性、准确度和 F1score。一些论文还报告了 AUROC（接收器工作特性下的面积）。超越上述指标有几个原因[98]。与医疗数据集相关的一个经常遇到的挑战是，它们在类别之间趋于不平衡，通常具有比病变图像更多的命名图像。因此，某些指标可能会对实际性能提供过于乐观的印象。例如，二元分类器可以通过将所有实例分配给正类，在包含少量负实例的数据集上实现高精度。众所周知，AUROC 对不平衡分类具有欺骗性[91]。在这种情况下，真实类和预测类之间的相关系数可以提供更多信息[15]，尽管没有一个指标是普遍信息或适合任何不平衡数据问题的。此外，出于检测目的，一个报告是每帧性能（即对视频中的每一帧做出决定）还是每帧（即对视频序列中的至少一个帧进行正确预测）也是一个问题。从技术角度看结果，通常需要每帧分析，但从医学角度来看，病变分析通常足以将发现通知临床医生一遍。

对于分段性能，常用的指标是 DSC 和 IoU，也称为 Jaccard 指数。在临床使用中，医学专家通常对有关潜在病变的像素详细信息感兴趣。DSC 和 IoU 可用于比较预测分割图与地面实况之间的像素相似性。此外，精度和灵敏度用于评估欠分割或过度分割，其中分割不足意味着与基本事实相比，模型预测图像某些部分的相关内容较少，以及过度分割，预测图像覆盖的像素多于基本事实。

如表 2 和表 3 所示，到目前为止，很少有研究关注所需的实时功能，以便在内窥镜检查期间向临床医生提供实时反馈。然而，似乎有报道的系统分析数据的速度超过了帧率阈值，并且在一些安排的比赛中也得到了关注（参见“比赛和挑战”部分）。尽管如此，这通常是速度（模型复杂性）和检测性能之间的权衡，这表明这仍然是未来 CAD 系统研究和开发中的一个重要问题。

自动生成报告

内镜医生完成内窥镜检查后，应生成高质量的报告。这通常是一个耗时的过程，研究表明，大约六分之一的美国医生工作时间花在行政任务上，占用了直接患者护理的时间并降低了工作满意度[115]。此外，内镜医生对研究结果的解释以及报告风格也存在很大差异。这可能导致最终决定的不一致[37]。因此，自动生成报告既可以节省临床时间，又有助于标准化内窥镜报告，自然语言处理的最新发展有望为自动报告生成开辟新的可能性[86]。

Jing 等人[51]提出的一种方法使用神经图像标题从 X 射线图像创建报告。在[121]中，图像通过神经网络进行分析，并且将类似于手头发现的示例图像和注意力图组合到报告中。大多数方法侧重于图像分析作为基础，并将其与其他信息相结合[24, 33, 118]。当然，这取决于对包含正确信息的数据库的访问，这些信息可以与图像结合使用。一个重大挑战是国家甚至医院之间的报告标准不同，这使得创建可广泛采用的软件几乎是不可能的。

然而，对于医学专家来说，自动文本创建甚至可能不是这种软件最关键的功能：一个更重要的方面是他们理解底层模型的推理和决策的能力，使他们能够将其包含在他们的评估中。这将在下一节中讨论。

可解释性

与基于深度学习的 CAD 系统相关的一个众所周知的挑战是，由于其固有的复杂性，可解释性有限[4]。这种特性使它们臭名昭著，成为决策过程未知的黑匣子，尤其是对最终用户而言[35]。需要理解和解释系统如何工作以及不同数据特征在决策中扮演的角色，以满足系统开发和使用不同阶段的不同需求。系统的开发人员需要了解数据和方法如何协同工作，因

为对输出的理解和可解释性有助于确定数据中的错误以及实现有针对性的故障分析。特别是，在这个目标的背景下，医学专家需要对系统的决定进行解释，以确保它与相关的医学知识一致。

基于深度学习的系统，如 CNN，没有提供解释的固有方式，这意味着它们必须扩展以包含解释生成器，或者必须在事后获得解释[1, 38, 39]。模型解释方法的简要概述如表 5 所示。模型可以设计为作为附加任务为其决策提供理由，例如，通过文本理由生成器作为模型架构的一部分[62]。给定一个没有这种设计的模型，可以使用不同的方法：那些解释决策系统本身的属性的方法，以及那些将系统视为黑匣子并根据其紧急行为提供解释的方法，分别称为模型依赖方法和模型不可知方法。前者的一个例子是将深度神经网络（DNN）的内部参数的值显示为叠加在分类实例上的热图[35]。如果解释正确，这可以提供对系统内部决策过程的理解。这种方法还可以扩展为包括关于系统诱导决策不确定性的信息（意思是不确定性的一部分与数据收集和选择过程无关），参见[113]。

在与模型无关的方法中，解释概念 LIME（局部可解释模型不可知解释）使用可解释模型（例如线性模型，决策树或下降规则列表）近似黑盒模型[84]。这是在实例的邻域中完成的，以便进行解释，使结果解释成为局部解释，因为它适用于单个结果并且基于特定实例的特征，就像上述依赖于模型的解释一样。

相比之下，全局解释捕获并解释整个模型，例如特征重要性排名。一类能够产生全局解释的方法是基于 Shapley 值的博弈论概念[92]，目前在统计学和机器学习文献中引起了极大的兴趣[27, 40, 44, 66]。Shapley 值是通过使用数据特征的所有可能组合评估模型获得的。因此，计算复杂度随着特征数量的增加而增加为 2，并且计算涉及为每个特征子集重新训练模型。后者是有问题的，因为重新训练会导致不同的模型参数，突出表明 Shapley 值只是模型不可知的，而不是独立的。广泛使用的 SHAP（SHapley Additive exPlanations）包[68]以不同的方式规避了各种模型架构的这些挑战，通过使用数据中的背景样本计算近似值，以及使用与 DeepLIFT 的每节点归因规则类似的方法进行深度架构[94]。Shapley 分解可以全局和局部计算，并且可以表述为 LIME 的特例[68]。Shapley 值也可用于获得与模型无关的解释[28]。

表 5 关于何时应用的不同模型解释方法：在模型开发期间（模型内）或模型完成后（模型后）。解释方法提供对本地（围绕特定预测）或全局模型行为的洞察

类别		描述	文献
模型内		作为模型体系结构一部分的对齐文本生成器	[62]
后模型	依赖模型	Grad Cam：在图像上显示 DNN 激活	[35]
	与模型无关	LIME：产生一个近似完整模型的局部可解释模型	[84]
		SHAP：全模型的条件期望函数的沙普利分解	[68]
	不依赖模型	全局非参数沙普利分解	[28]

比赛和挑战

与内窥镜数据的自动分析相关的一系列不同挑战[9, 36, 79]，其中基于 CNN 的方法在过去几年中一直是表现最好的方法。给出的各种任务是基准测试和开发自动化系统，以准确检测、定位和分割胃肠道内的异常。这些挑战针对不同的任务，从胃肠道异常、结直肠息肉的检测、定位和分割到胃肠道中存在的伪影（见表 6）。这些定期竞赛可以帮助该领域的研究界找到评估模型的共同标准，对最先进的方法和工具进行基准测试，并找到新的方向，使该领域共同向前发展。

表 6 GI 检测、分类和分割挑战示例列表

赛事名称	链接
MICCAI 2015 Endoscopic Vision	https://polyp.grand-challenge.org/databases/
Medico 2017	http://www.multimediaeval.org/mediaeval2017/medico/
Medico 2018	http://www.multimediaeval.org/mediaeval2018/medico/
GIANA 2018	https://giana.grand-challenge.org/Home/
EAD 2019	https://ead2019.grand-challenge.org/
Bio media 2019	https://github.com/kelkalotbiomedia-2019
Medico 2020	https://multimediaeval.github.io/editions/2020/tasks/medico/
Endo Tect 2020	https://github.com/simula/icpr-endotect-2020
EDD Challenge 2020	https://edd2020.grand-challenge.org/
End oCV 2020	https://endocv.grand-challenge.org

表 7 新兴商业产品

产品	供应商	年份	网址
GI Genius A I	Medtronic/Co smo Pharma	2019	https://www.cosmopharma.com/products/gi-genius
Endo BRAIN- EYE	Cybernet	2020	<a href="https://www.cyberetjp/english/documents/pdf/news/pr
ess/2020/20200129.pdf">https://www.cyberetjp/english/documents/pdf/news/pr ess/2020/20200129.pdf
CAD-Eye	Fujifilm	2020	https://www.fujifilm.eu/eu/cadeye
Ai4Gi	Ai4Gi	2016	https://ai4gi.com
Ulti Vision	Do cBot	2018	https://www.docbot.co/gastroenterology-and-health
DISCOV ERY	Pentax	2020	<a href="https://www.pentaxmedical.com/pentax/en/95/2/DIS
COVERY-new">https://www.pentaxmedical.com/pentax/en/95/2/DIS COVERY-new
ENDO- AID	Olympus	2020	<a href="https://www.olympus.no/medical/en/Products-and-
Solutions/Products/Product/ENDO-AID.html">https://www.olympus.no/medical/en/Products-and- Solutions/Products/Product/ENDO-AID.html
SOMA	Auger e Medical	2018	https://augere.md

临床验证和新兴商业系统

许多研究小组已经提出了有希望的研究成果和良好的性能指标，并且出现了几种基于 AI 的商业系统，其中一些列在表 7 中。这些的状态大多是未知的，但是，例如，GI Genius 系统带有 CE 标志，但仍然缺乏美国食品和药物管理局（FDA）的批准，而 EndoBRAINEYE 仅在日本获得批准。对于部署 CAD 系统用于临床检查室的实时检查，或用于 VCE 数据后分析，临床验证是绝对必要的。尽管如此，在撰写本文时，此类研究非常有限。2020 年 8 月，Repici 等人 [83] 提出了一项随机多中心试验，得出的结论是，基于 AI 的 CAD 提高了腺瘤检测率（ADR），即至少有一个组织学证明的腺瘤或癌的患者百分比，证明了这种系

统的潜力。他们检查了 685 名患者：341 名患者使用 CAD 系统，344 名患者仅使用传统的手动检查。该系统的 ADR 为 54.8%，对照组 40.4%。这表明基于人工智能的系统可以帮助检测腺瘤，但需要进一步改进以提高检测率，并检测更多的无柄锯齿状病变（根本没有）。考虑到研究的局限性以及呈现的性能，很明显仍有改进之处，并且需要进行更多的临床研究。

尽管业界对此非常感兴趣，但普遍缺乏有关评估方法和可重复性的适当标准。此外，行业应用似乎并不关注模型可解释性或模型输出可解释性。这些都是值得信赖的应用的关键要素，行业发展有望遵循当前的研究趋势，并在未来更多地关注这些。

最后，当高性能（研究）原型已经构建和测试，满足上述要求时，必须批准用于医疗用途。在实施之前，需要对基于人工智能的软件进行稳健的评估，以降低患者和卫生系统的风险，建立信任以促进广泛采用。此类产品的常用术语是基于 AI 的软件作为医疗设备(SaMD)。SaMD 申请的监管机构，包括美国的 FDA，一直受到全球协调工作组和国际医疗器械监管机构论坛（IMDRF）的指导。IMDRF 为 SaMD 提出了四种不同的风险类别，每种类别都有一套不同的要求来评估该技术的科学和临床有效性[42]。在胃肠病学中，CADe 和 CADx 技术尚未分类。目前 FDA 对 SaMD 的审批流程源自其医疗器械审批流程，将分为 I 类、II 类和 III 类（最高风险）[105]。风险分类后，作为 510 (k) 途径或从头途径提交的上市前文件可能与类似于 Osteoidetect 的基于 GI 的 AI 技术相关[107]。此外，鉴于人工智能算法是快速迭代和不断学习的，它可能对当前的监管过程构成挑战。FDA 在其数字健康创新行动计划中提出了针对 AI 技术的新监管体系，重点关注依赖于持续学习和适应的 AI 技术[106]。世界各地的监管机构也认识到人工智能算法应用于医学时所面临的挑战，大多数国家已经开始努力制定针对 SaMD 的政策。他们中的许多人共享指定风险的核心原则，审查临床证据以证明有效性和安全性，以及纳入不断发展的人工智能系统的实践。

总结和结论

在这项工作中，我们介绍了自动数据分析在胃肠道内窥镜检查中的应用，并概述了基于检测和分割的方法，以应对大病变漏诊率和观察者间变异性等挑战。最近的研究表明，基于深度计算机视觉的方法似乎有可能通过提供全自动 CAD 系统作为额外的数字眼睛来提高胃肠道内窥镜检查的准确性和整体性能。然而，在自动分析可以有效地整合到临床实践之前，仍有一些悬而未决的问题和挑战需要解决。这些应被视为需要在该领域进行研究关注的问题。

致谢

这项工作部分由挪威研究委员会资助，项目编号 282315 (AutoCap)。

摘要

人工智能（AI）正在彻底改变医学实践的方式。在这种情况下，人工智能算法在内窥镜中的应用越来越受到关注，因此调制解调器内窥镜正朝着越来越多的辅助/自动解决方案发展。为了提高诊断和外科手术的准确性，已经采取了几种方法。本章概述了该领域的主要贡献。确定了四大类应用，即：(i) 内窥镜手术过程中的检测和诊断，(ii) 信息框架选择，(iii) 镶嵌和表面重建，(iv) 用于术中辅助和外科医生培训的增强现实系统。对未来研究方向和临床实践中的实施进行了讨论。

关键词

内窥镜检查 人工智能 机器学习 深度学习 计算机辅助检测 ·计算机辅助诊断 信息帧选择
马赛克 增强现实

导论

内窥镜检查是一种用于检查人体内部器官的微创程序：将带有灯和摄像头的柔性管插入患者体内以观察内部器官和组织的细节（图 1）。内窥镜检查也可用于执行其他任务，包括小手术。为此，内窥镜可以通过所谓的工作通道配备手术工具。诊断小肠特定疾病的体格检查也可以通过胶囊内窥镜（CE）进行，CE 是一种嵌入摄像头的药丸形装置，可穿过胃肠道，捕获图像并将其传输到外部接收器[1]。当胶囊穿过消化道时，它会拍摄数千张照片，这些照片被传输到外部接收器。

内窥镜检查可用于检查许多解剖区域：呼吸道，包括鼻子（鼻镜检查）和下呼吸道（支气管镜）；尿路（膀胱镜检查）；食道和胃（胃镜检查）；结肠（结肠镜检查）；腹腔或盆腔（腹腔镜）、关节内部（关节镜）、胸部器官（胸腔镜和纵隔镜）[2]。调制解调器内窥镜检查程序起源于 50 年代后期，已成为诊断和治疗人体内各种疾病和损伤的最重要方法[3]。内窥镜检查通常被认为是一种相对安全的手术，通常在患者有意识时进行。然而，这对外科医生来说可能是一种负担，其表现可能会因疲劳、压力或经验有限而改变[4]。随着人工智能（AI）在医学中的兴起，包括计算机视觉的最新进展，已经进行了许多研究人员以提高该领域的性能，因此现代内窥镜检查正朝着越来越多的辅助/自动程序发展。

人工智能辅助内窥镜检查基于模仿人类认知功能的计算机算法。该想法基于算法合理化和采取最有可能实现特定目标的行动的能力。特别是，在内窥镜领域成功应用的各种人工智能技术中，最重要的成就可归因于机器学习（ML）。ML 是一类用于识别数据模式的通用算法。ML 算法无需显式编程即可从经验中学习，从而在试错迭代过程中提高其能力。在这种情况下，一个重大突破是深度学习（DL），这是机器学习算法的一个快速增长的子集，受人脑的生物学启发。特别是，卷积神经网络（CNN）已被证明非常适合图像和视频处理任务。CNN 利用多个网络层（对图像补丁执行的连续卷积操作）从图像中提取关键特征，并通过全连接层提供最终分类作为输出。这样一个相对简单的想法使得在计算机视觉方面取得了令人印象深刻的结果，包括与内窥镜相关的任务。

人工智能在内窥镜检查中的几种应用可以在文献中找到。概述如图 2 所示。

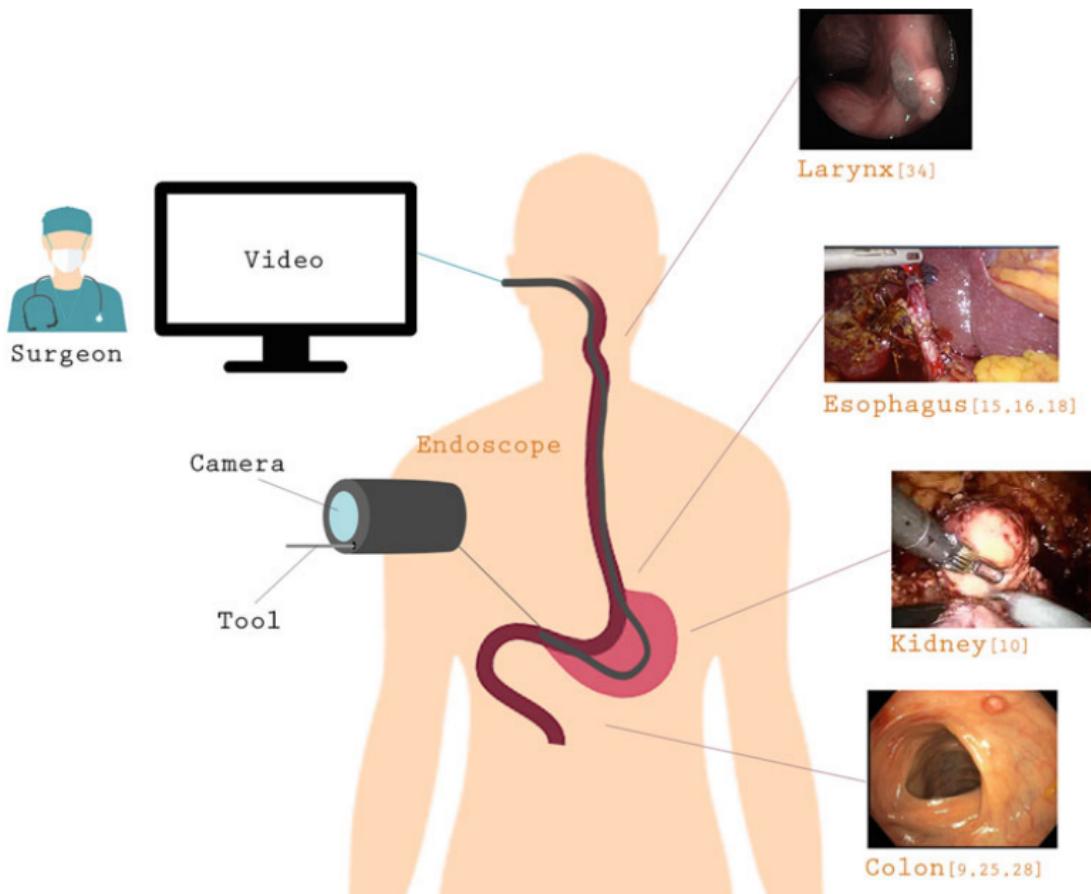


图 1 内窥镜检查程序：将内窥镜插入患者体内，用于观察内部器官和组织的细节。它也可以用来执行其他任务，包括小手术

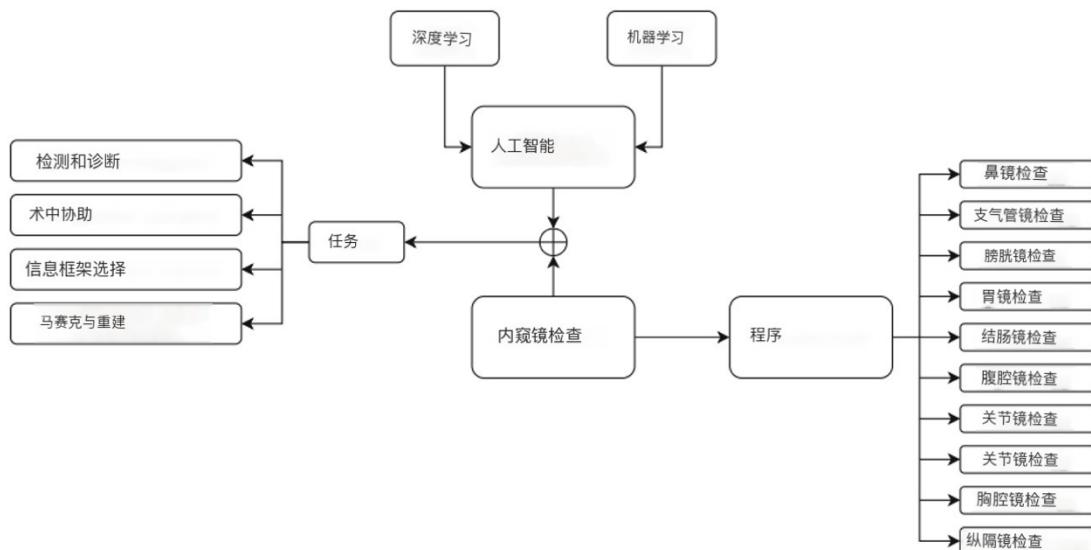


图 2 人工智能在内窥镜中的应用综述

在最有效的应用中，检测和诊断任务是研究最多的任务之一。特别是，研究兴趣似乎主要与胃肠道息肉及其检测和分类、病变模式识别、良性和癌前息肉分化是分布最广泛的应用

有关[5]。通过天真地在 Pubmed 上搜索“人工智能”、“机器学习”、“深度学习”和“内窥镜”等词，确实，最近 20 年增加了大约 220 个结果，其中 70 个与“诊断”，“检测”和“表征”有关，超过 80 项研究调查胃肠道。但是，快速增长和里程碑应用程序仍然存在。这包括计算机辅助系统的实时应用、视频内窥镜检查中的信息帧选择、镶嵌和表面重建。

在本章中，将详细研究这些贡献。将提供研究最充分的任务的概述，重点是经典方法和最新趋势。将分析现有的挑战，以收集对未来工作的见解。

人工智能在内窥镜实践中的应用

在下文中，将提供与内窥镜中 AI 相关的主要主题的一般概述。

内窥镜手术中的检测和诊断

人工智能的医疗应用在过去几年中取得了重大进展。在微创手术（MIS）领域，计算机视觉系统已成为当今的热门话题。在可能的任务中，计算机辅助检测（CADe）和计算机辅助诊断（CADx）系统受到了外科数据科学（SDS）社区的特别关注[3]。CADe 系统有助于定位视野内的元素。并且已经提出了几种技术来帮助内窥镜医生在实时和离线应用中标记显眼的结构和部分。通常，这是通过在图像上围绕感兴趣区域放置一个边界框来实现的。或者，可以在描绘目标元素轮廓的地方执行分割任务[6]。最终，CADx 系统技术不仅检测感兴趣的区域，还用于在标准筛查和监测期间表征和分类元素（例如，评估特定形式的癌症的分期或诊断某些感染）[4]。从 90 年代开始，已经应用了几种技术，最初基于图像处理技术（区域增长，阈值等）和手工制作的特征[7, 8]，然后与更复杂的 ML 算法相结合以提高性能[9]。然而，随着 DL 的兴起，DL 取得了重大突破，可以从经典的内窥镜图像中自动提取特征，而无需放大窄带成像（NBI）、柔性光谱成像颜色增强（FICE）或蓝色激光成像（BLI），这些在临幊上可用于区分癌区和非癌区[10]。已经提出了单帧分析或视频序列处理方法。研究最多的应用之一是息肉的计算机辅助检测和计算机辅助诊断[11]。息肉是从粘膜突出的组织异常生长。结肠息肉是最常见的，但也有可能在其他解剖区域（包括胃和子宫）发展息肉。尽管大多数息肉是无害的（良性的），但其中一些最终会发展成癌症。在过去的 20 年中，自动息肉检测和分类一直是一个活跃的研究课题，并且已经提出了不同的方法[12, 13]。然而，患者常规治疗的应用非常有限[9]。实时约束、息肉外观的巨大变化、褶皱、血管和管腔等其他元素的存在会影响性能，并使 CADx 和 CADe 系统的开发变得非常具有挑战性[9]。

人工智能还被用于开发与身体多个部位的几种疾病相关的诊断和检测系统，包括食道、胃、小肠和结肠。例如，与巴雷特食管（BE）相关的早期异型增生的光学诊断目前只能由专家精确完成[10]。巴雷特食管是食管腺癌发展的决定性因素，自动识别 BE 中异常增生变化是研究 CADe 胃镜的流行趋势之一。尽管如此，已经开发了几种基于 AI 的技术，取得了可喜的结果[12, 14]。在这种情况下，早期的工作考虑了外观特征（颜色，脉管系统和表面图案）来训练经典分类器，如支持向量机（SVM）[15]。Van der Sommen 等人[16]提出了一种基于特定纹理、滤色片和机器学习的肿瘤检测方法。该研究考虑了来自 44 名患者的 100 张图像，灵敏度和特异性为 83%。为了达到同样的目标，Swager 等人 [17] 在 19 名患者的 60 次近显微分辨率扫描上测试了一个 AI 模型。结果显示敏感性和特异性分别为 90% 和 93%。最近，已经提出了先进的 DL 技术，其目标是提高性能并避免手动指定特征的耗时步骤[15]。De Groof 等人[18]提出了一种适合在临床实践中实时使用的 DLCADe 系统，以改善 BE 患者早期肿瘤的内镜检测。该方法使用来自 669 名患者的 1704 张独特的食管高分辨率图像进行了测试，实现了 90% 以上的准确率。

SCC 的鉴定也受到特别关注[14]。鳞状细胞癌是全球最常见的食管癌，最常见于食管上

部和中部[19]。Guo 等人开发了一种用于诊断癌前食管病变的实时 DL 系统[20]。该模型创建一个概率图，指示疑似肿瘤区域和非癌性区域。它使用 6000 多张 NBI 图像进行了测试，灵敏度为 98.04%。Ohmori 等人使用神经网络进行食管病变检测和分化[21]。使用了 10,000 多张内镜非放大图像，灵敏度、特异性和准确度分别为 90%、76% 和 81%。人工智能可以为胃病的管理提供有价值的帮助。在这些应用中，胃癌的检测和幽门螺杆菌 (HP) 感染的识别被广泛研究。HP 相关性慢性胃炎可引起黏膜萎缩和肠上皮化生，这两者都会增加胃癌的风险[22]。在 Itoh 等人[22]中，CNN 用于准确诊断 HP 感染。该系统使用从 139 名患者获得的 179 张上消化道内窥镜图像进行测试。CNN 检测 HP 感染的敏感性和特异性为 86.7% 和 86.7%，中华民国下面积为 0.95。一些研究还证明了人工智能在无线胶囊内窥镜检查 (WCE) 中的应用。WCE 是一种用于小肠内窥镜探索的技术[23, 24]，WCE 被证明有用的病例在过去几年中显着增加。WCE 对临床医生来说是一个主要挑战，即审查图像的时间紧迫性[25]。出于这个原因，ML，特别是 DL，代表了一个非凡的观点，并有助于扩大适用范围。根据最近的相关综述，ML 已被广泛用于增强 WCE 分析，包括消化道出血的检测、小肠溃疡和糜烂的检测、胃肠道血管扩张的检测等[23]。

最后，CAD 系统，特别是 CADe，不仅限于疾病检测和表征。考虑到传统内窥镜检查中可能存在的元素数量，一些研究人员开始开发检测和跟踪解剖外其他物体的系统[26]。仪器检测、分割、跟踪和姿态估计是这方面的一些例子。然而，光照水平、背景变化和视野中工具数量的不同给算法和模型训练带来了困难[27]。

信息丰富的帧选择

人工智能有助于改善或监测内窥镜质量。在这种情况下，已经实施了信息帧选择系统来监测质量标准并减少要处理的信息量，例如，指示可能导致诊断错误的退化因素，例如噪声、采集错误、眩光、模糊和照明不均匀[28]。手动修改内窥镜视频确实既昂贵又耗时。例如，在 WCE 中，典型的记录由超过 50,000 帧组成，需要大约 4590 分钟来审查它[29]。为了尽量减少这些数据量，一个可能的步骤是消除没有诊断相关信息的帧，这只会降低视频分析的准确性[30]。这些解决方案已经对临床结局产生了更大的影响，因此可能越来越成为内镜检查感兴趣的领域[31]。

在 Alizadeh 等人[32]中，提出了一种多阶段方法，包括主动轮廓，颜色范围比，自适应伽马校正方法，精明的颜色边缘检测运算符和形态学处理[33]，以区分内部粘膜（信息性）和非信息区域（灵敏度和特异性为 81% 和 92%）。在 Moccia 等人[34]中，一组强度，基于关键点和纹理特征以及多类 SVM 用于对 NBI 喉镜视频中的信息性和三类无信息帧进行分类（在来自 18 个不同喉镜视频的 720 张图像的平衡集上测试的平均 AUC 为 91%）。最近，已经表明深度学习算法可能优于图像分析的标准学习方法。Patrini 等人[35]提出了一种在喉镜中选择信息框架的新方法，也显示了该领域迁移学习的潜力。SVM 和基于 CNN 的方法用于将帧分类为信息性和无信息性帧，例如模糊、唾液或镜面反射以及曝光不足。该系统在 18 个 NBI 内窥镜视频的数据集上进行了测试，参考了 18 名受鳞状细胞癌影响的不同患者，实现了 0.98 的 ROC 曲线下面积。

镶嵌和表面重建

内窥镜手术的优点是提供微创干预，这意味着患者和外科医生的负担减轻，以引导内窥镜并了解其在体内的位置。然而，对于后者，这种技术需要高度的协调性和精细运动能力来引导内窥镜并意识到其在体内的位置。原因包括内窥镜提供的视野非常有限，以及显示的图

像与物理环境之间没有直接关系[2]。这些问题提高了研究人员和行业的兴趣，即通过创建增强的视野来提供辅助内窥镜医生的方法。

镶嵌（也称为拼接）是指将内窥镜图像视频序列中的多个（部分重叠）帧组合在一起以创建具有更宽视角的更宽视野或全景图像的过程。在典型的管道中（图 3），在内窥镜手术期间采集图像并进行预处理，以清理和减少输入数据量（例如，对视频流中可能包含冗余信息的后续图像数量进行子采样）[36, 37]。

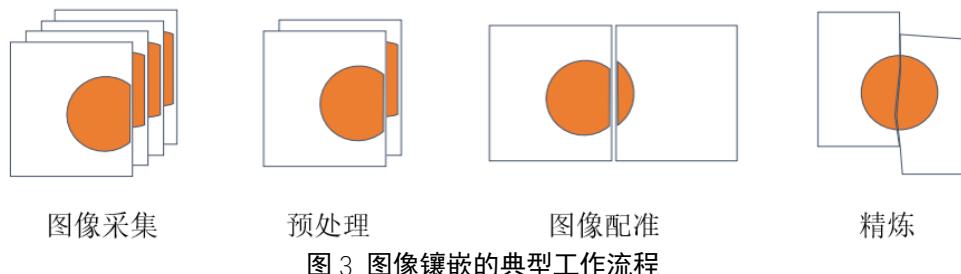


图 3 图像镶嵌的典型工作流程

然后，执行图像配准，这是对齐代表物理连接组件的图像的任务[2]。常见的方法包括基于像素或基于特征的对齐方法。前者考虑重叠区域内的整个像素组，而后者提取更高级别的特征并利用其相似性来计算匹配[38, 39]。通常，对获得的重建执行进一步的优化过程，以提高最终结果的准确性[2]。最初由摄影和摄影测量界开发的几种技术已经应用于每个步骤。在 Szeliski [40], Bergen 和 Wittenberg[2]提供了全面的调查。

可以通过将图像投影到平面上来应用 Mosaiking，或者，如果图像中观察到的场景具有显着的 3D 结构并且相机执行任何平移运动，则可以将其投影到更高维的空间上。在这里，镶嵌和三维表面重建相交，可以认为涉及类似的问题。在这种情况下，现有工作提出了一种软件解决方案，以使用内窥镜视频中估计的内窥镜姿势重建目标器官（例如结肠，肝脏和喉部）的 3D 结构。这些方法包括阴影形状 (Sfs)，运动结构 (SfM)，视觉同步定位和映射 (SLAM) [41, 42]。SF 试图通过考虑照明和表面反射率的变化来检索人体组织表面的 3D 结构。尽管其假设很简单（例如，所有可见表面点将从同一方向接收相同数量的光[43]），但这种方法在内窥镜应用中提供了有希望的结果[44, 45]。但是，单独的 Sfs 技术可能不适用于在操作过程中照明条件急剧变化的场景。出于这个原因，研究人员已经开始将 Sfs 方法与 SfM 方法相结合，同时利用阴影和特征信息从内窥镜图像重建表面[46]，并与 SLAM 一起实时利用内窥镜轨迹[47, 48]。Bergen 等人[2]详细介绍了用于镶嵌和表面重建的里程碑技术，并根据医学分支和解剖结构调查了相关应用。最近，深度学习技术已经开始引起研究人员的兴趣。Turan 等人使用 DL 到 3D 表面重建方法进行内窥镜胶囊机器人[49]。从 2D 图像预测内窥镜深度信息的任务引起了特别的兴趣，其中已经提出了几种深度学习方法，并取得了有希望的结果[50, 52]。然而，已经进行了一些关于单应性估计[53, 54]和胎儿镜镶嵌[55]的工作。

用于术中辅助和外科医生培训的增强现实系统

内窥镜模拟器用于虚拟再现现实生活中的临床环境。这一点尤其重要，因为外科医生在手术室的身体和精神负担都可以减轻，受训者可以通过直接在现场患者身上练习来获得技术熟练程度[56]。增强现实 (AR)、虚拟现实 (VR) 系统提供了几种可能性：一方面，外科医生可以在手术过程中利用增强的视野。另一方面，它们允许受训者通过在虚拟患者身上练习来体验模拟临床场景，测试他们的反应并从 CAD 系统实时受益[57]。在 AR 中，通过将实时视频输入与沉浸式显示技术（例如，计算机生成的 3D 对象叠加在视频帧上）相结合，增强

了用户的真实世界。目的是在观察患者时向外科医生提供有关患者更多的信息。这与 VR 不同，VR 中环境的基本元素完全由计算机模拟，以模拟它们的存在[58]。尽管 AR 的历史悠久，可以追溯到 60 年代，但直到最近几年，AR 才开始用于医疗实践。人体内特定解剖结构的定位是首批应用之一[58, 59]。Bertrand 等人[60]提出了一种技术，该技术使用一种称为 Hepataug 的新软件将可变形的术前模型半自动地叠加到腹腔镜图像上。他们的分析报告了在腹腔镜肝切除术中使用 Hepataug 通过可变形模型实现 AR 的可行性和潜在兴趣。Hussain 等人[61]开发并评估了基于视频的增强现实系统的性能，该系统结合了术前计算机断层扫描和实时显微视频。将虚拟内窥镜图像配准到基于基准标记的完整鼓膜的显微镜视频中，并在显微镜运动过程中应用单应性变换。特别是，该系统提供了有关中耳结构和手术器械的额外视觉信息，精度为亚毫米级，与中耳手术兼容。然而，自从引入可穿戴平视显示器以来，外科界对将这项技术应用于常规外科实践非常感兴趣。在 Yoon 等人最近的一项调查中[62]，在 74 篇已发表的评估可穿戴抬头显示器在手术环境中的效用的文章中，其中许多与内窥镜外科手术有关。最近，在 AI Janabi 等人[63]中，Microsoft HoloLens 已被评估为传统内窥镜监视器的可行替代方法，表明该设备有助于改善性能结果，并被研究参与者广泛接受为手术视觉辅助工具。在 Qian 等人[64]中，作者介绍了 ARssist，一种基于光学透视头戴式显示器的增强现实应用程序，以帮助第一助手更有效地执行他/她的任务，从而改善机器人辅助腹腔镜手术的结果。

讨论与展望

人工智能在内窥镜中的使用越来越受到关注，因为它有可能在许多层面上提高内窥镜检查的质量。内窥镜手术过程中的实时和离线检测和诊断取得了可喜的结果。然而，在它们在临床应用和医疗常规中得到有效应用之前，仍然存在各种问题和挑战。第一个主要限制在很大程度上依赖于算法泛化能力。正如在最近的相关工作中已经指出的那样[11, 12, 23]，应进行多中心研究，以避免与用于获取数据的机器相关的任何偏差。此外，验证还应侧重于非平凡的实例（例如，具有低可变性的高质量帧或序列），以免高估性能。在这种情况下，应在论文中提及该研究的局限性，并应纳入在临床环境中有效使用的建议，以帮助理解未来的步骤[31]。此外，应提高模型（尤其是 DL）的可解释性，以提高算法的可靠性。尽管如此，一些人工智能产品已经获得监管部门的批准，并可能在未来用于临床。例如，支持结直肠息肉诊断的 AI 系统 EndoBRAIN（日本东京奥林巴斯）已获得监管部门批准，现在与内窥细胞镜系统联合提供，因此有助于组织的实时体内表征[65]。深度学习解决方案（GI Genius；美敦力，明尼阿波利斯，明尼苏达州，美国）已经发布，用于提高回肠结肠镜检查息肉检出率[25]。因此，CAD 系统可以代表内窥镜下一个未来的重要支持。新颖和有前途的工具即将开发出来。然而，应向内镜医生提供仔细的指南，以帮助他们了解此类新技术的潜力[31, 66, 67]。这对于确定内窥镜手术中检测和诊断的未来至关重要。此外，内窥镜外科实践中的 AR 是一个有价值的研究领域，可能会在未来显著提高手术质量。可穿戴平视显示器的使用值得研究，因为与传统手术相比，它们通常可以改善人体工程学，在传统手术中，外科医生需要反复转动头部以在手术区域和影像监视器之间转移视觉焦点[62]。在这个方向上，减小屏幕尺寸以及全息投影代表了改进的领域。关于镶嵌和表面重建，仍然可以观察到一些挑战。首先，缩短实时应用程序的算法处理时间是一个重要目标。一些应用在提供用于临床环境的全景视图方面仍然取得了有价值的成果，而加速到实时能力仍然是一个问题[2]。最后，在非刚性结构的重建和镶嵌方面成功开展的工作很少[68, 69]，取得了有价值的成果。然而，这项任务应该考虑改进的空间和新的想法，这可以被视为尚未解决的挑战[41]。需要一个处理呼吸变形、心脏运动、器官移位和组织工具相互作用的框架。为此，DL 解决方案可以被认为是一个有

价值的领域，可以探索可以在模拟变形模型时训练模型的地方[70, 71]。

结论

总之，本章概述了研发的现状以及人工智能在内窥镜中的应用前景。这些技术开辟了新的可能性，并且可能通过帮助医生而不是降低其相关性而在临床实践中变得有效。

P951-P957 (BE Cancer Detection Using NarrowBand Imaging 之前)

巴雷特食管中的 AIM

Joost van der Putten and Fons van der Sommen

摘要

在过去十年中，胃肠病学的成像质量急剧提高。目前最先进的内窥镜设备能够通过白光内窥镜 (WLE) 和几种光学色素检查技术捕获高清镜头。这种强大的改进导致了内窥镜领域从可视化到解释的范式转变。与此同时，计算机科学领域已经快速发展到可以使用当前的计算硬件来开发复杂的人工智能 (AI) 系统，用于各种应用，包括内窥镜成像。巴雷特癌症的预后在很大程度上取决于诊断时的分期。不幸的是，大多数癌症仅在晚期发现。由于大多数内窥镜检查是在外围医院进行的，因此对视觉表示有特定的专业知识和熟悉。

关键词

巴雷特食道 人工智能 深度学习

导论

问题陈述

食管癌是全球第八大常见癌症，也是癌症相关死亡的第六大原因[1]。尽管多年来内镜的图像质量有了很大改善，但当由必须承担大部分监测的普通内镜医生进行监测时，早期病变的检测经常失败[2]。此外，80%的病例发生在欠发达地区，这些地区获得经验丰富的医生和专业设备的机会较少。这些因素突出表明需要一种可靠、廉价和通用的解决方案来检测食道中的癌症。

大多数食管癌可细分为两种主要的组织学亚型：食管腺癌 (EAC) 和食管鳞状细胞癌 (ESCC)。一种称为巴雷特食管 (BE) 的预先存在的疾病通常是 EACS 的起源。当食道下部的正常粘膜细胞被通常只存在于小肠和大肠中的抗酸细胞取代时，BE 在食管的下三分之一发生。相反，鳞状细胞癌主要发生在食管上三分之二的扁平细胞中[3]。不幸的是，EAC 和 ESCC 的预后都很差，因为病变主要在晚期发现。早期发现至关重要，因为随机试验表明，低发病率的内镜治疗（例如，通过消融或切除术内镜根除异型增生）可显著降低晚期癌症的发生风险[4, 5]。在早期（即仅限于黏膜或肌层黏膜）内镜下切除食管癌可达到>90%的治愈率，而无需行食管切除术，食管切除术的发病率很高[6, 7]。

食道中的人工智能案例

在过去的几十年里，胃肠病学的成像质量急剧提高。目前最先进的内窥镜设备能够通过白光内窥镜 (WLE) 和几种光学色素检查技术捕获高清镜头。这导致了内窥镜领域从可视化到强调解释的范式转变。与此同时，计算机科学领域已经快速发展到可以使用当前的计算硬件来开发用于各种应用的复杂 AI 系统[8, 9]，包括内窥镜成像[10, 11]。这两个因素相结合，在食道中有效使用人工智能铺平了道路。特别是深度学习领域的进步对此类系统的数量产生了巨大影响[12]。

正确设计和测试的人工智能系统对于内窥镜领域非常有价值。最明显的用例是癌症的早期检测。EAC 和 SCC 都是特别致命的疾病，因为许多早期病变被漏诊，然后发展为无法再内镜治疗的癌症[13]。尽管治疗取得了进展，但诊断为食管癌的患者的预后仍然较差，基于人群的研究显示 5 年生存率为<20%[14, 15]。食管癌的预后在很大程度上取决于诊断时的分期。不幸的是，大多数癌症仅在晚期发现[16, 17]。由于大多数内窥镜检查是在外围医院进行的，在这些医院中，通常不存在对早期病变视觉表示的特定专业知识和熟悉程度，因此当人工智能的性能与专家相当时，计算机辅助检测系统可以挽救无数生命。食道早期癌症的视觉识别非常困难，是一项高度主观的任务，需要大量的努力和经验。即使是专家，使用最常见的 WLE 方法的灵敏度也仅为 48%[18]。因此，引入了西雅图方案[19]，其中在四个象限中每 12cm 进行一次活检。然而，在实践中，该方案的依从性较低，研究发现内镜医生系统性地对长段 BE 患者样本不足[20]。这会导致更多的病变漏诊和癌症进展。活检指示剂可以显著缓解这些问题。

AI 系统的有效性高度依赖于内窥镜检查的质量[21]。一个给定的具有完美精度的计算机辅助系统理论上可以解决许多问题。然而，如果病变没有正确成像，这些系统就不能充分发挥其潜力。因此，患者也可以从监视内窥镜检查的自动质量检查中受益匪浅。患者通常仅使用部分镇静剂，因此食管在常规内窥镜检查期间不会受到阻碍，尤其是在没有经验的医生进行手术时，患者反射会引起宫缩，这可能导致漏诊病变。在这种情况下，人工智能可以通过提供有关视频质量的反馈或自动选择最佳帧进行进一步分析来帮助医生。虽然人工智能在许多研究和临床领域（包括胃肠道成像）呈上升趋势，但仍存在许多挑战。这些类型研究的技术和医学方面的同时整合通常很困难，并且这些系统的物理实施比其他不受隐私法约束的领域要费力得多[22]。这项工作试图展示巴雷特食道中与人工智能有关的研究，这将有助于弥合这一差距。

巴雷特食道的人工智能

巴雷特食管 (BE) 是一种医学疾病，其中食道下部的正常粘膜细胞被通常仅存在于小肠和大肠中的抗酸细胞所取代。BE 被认为是一种癌前病变，因为该疾病与通常致命的癌症食管腺癌 (EAC) 的高发病率有关。患有 BE 的患者发生 EAC 的变化增加了 30 倍。如果早期发现 BE 癌症，仍可进行内镜治疗，且预后良好[23]。然而，在许多情况下，EAC 的早期阶段未被发现，因为常规内镜操作由不熟悉疾病早期视觉体征的内镜医生进行[24]。因此，推荐的方案需要在整个食管中进行均匀间隔的活检。不幸的是，研究表明，目前的活检方案容易出现采样错误，从而导致遗漏病变。因此，相对容易治疗的早期癌症发展为晚期癌症，其中侵入性手术和放化疗是唯一可用的治疗选择。在以下部分中，我们将讨论人工智能系统在巴雷特食道中的使用。相关参考文献的简要概述如表 1 所示。

使用白光内窥镜检测 BE 癌症

在深度学习革命之前，许多研究人员试图使用传统的机器学习技术对巴雷特食管中的异型增生进行分类。Van der Sommen 等人[25]使用 Gabor 滤波器从概述中的 BE 图像补丁中提取特征。然后，这些特征用于训练支持向量机 (SVM) 分类器，以对新补丁进行预测。然后可以通过组合补丁分数来创建分段。同样，Souza 等人[26]也使用 SVM 对 BE 中的异型增生进行分类，但使用颜色重合矩阵提取特征。其他人则试图通过将深度学习和手工制作的特征方法合并到同一模型中来弥合两者之间的差距。在 Ghatwary 等人的工作[27]中，Gabor 特征与从 DenseNet 架构[28]中提取的特征相结合，对 Kvasir 数据集[29]和来自内窥镜视觉挑战的

数据集[30]进行分类。Van Riel 等人[31]确实使用了 CNN，但使用了所谓的 CNN 代码。通过这种方法，使用预训练的 CNN（通常使用 ImageNet 数据集进行预训练），其中特征是从倒数第二层中提取的。在这项工作中，Alexnet [32]，VGG [33] 和 GoogleNet [34] 被用作特征提取器，SVM 和随机森林[35]用于进一步分类图像。Mendel 等人发表了使用 CNN 检测食道腺癌的第一个初步结果[36]。在这项工作中，通过在全分辨率图像的补丁上微调 ResNet 模型 [37] 来使用相同的内窥镜视觉挑战数据集。从那时起，同一小组使用深度学习对 WLE 图像进行了更多研究。在他们最近的深度学习 WLE 研究中，Ebigbo 等人[38]开发了一种具有更先进的深度学习技术的新算法，例如空间金字塔池[39]和最先进的解码器[40]。值得注意的是，该算法不是在概述的 WLE 图像上进行训练和测试，而是在放大的病变/健康组织上进行训练和测试。因此，该算法不能用于初级检测，但系统表现出优异的性能，总体精度为 89.9%。在该小组最近的工作中，建议使用生成对抗网络(GAN)[41]来增加数据增强的变化[42]。巴雷特食管的假图像与真实示例一起生成，以增加数据集的变化。结果表明，与使用原始数据集相比，生成 BE 的人工补丁可以提高分类性能。另一个小组对许多现有的深度学习架构进行了广泛的评估，以检测早期食管腺癌。在 Ghatwary 等人的工作[43]中，几个对象检测框架，如基于区域的卷积神经网络[44] (RCNN)，快速 RCNN[45]，更快的 RCNN[46] 和单次多盒检测器 (SSD) [47] 被实施和评估来自内窥镜视觉挑战的数据。这项工作的结果表明，即使数据有限，深度学习在检测癌症方面也可以胜过传统的机器学习方法。同一小组最近还提出了一种从内镜视频中自动检测内镜异常的系统[48]。在这项工作中，将 3D CNN 与卷积长短期记忆 (LSTM) 相结合，以有效地学习短期和长期时空特征。然后，区域提案网络利用学习到的特征图来生成在视频中指示癌症的边界框。与其他深度学习研究相比，上述关于 BE 内窥镜癌症检测的出版物使用的数据相对较少，而其他深度学习研究通常使用更多数据表现更好。2020 年，Hashimoto 等人 [49] 发表了一项研究，与其他只能访问较小数据集(数百张)的小组的早期研究相比，其中包含更多的数据(近 1000 张图像)。数据集包含 WLE 和 NBI 图像(包括缩放图像)。CNN 算法经过预训练(在 ImageNet 上)，随后进行微调以执行二进制分类。一个额外的对象在算法上检测，目的是在发育不良区域周围绘制边界框。令人印象深刻的结果是分类准确率为 95%，平均准确率为 0.7533 用于检测的是实时性能。然而，该算法是在有限数量的患者(总共 100 名)上进行训练的。De Groof 等人[50]和 Van der Putten 等人[51]最近证明了使用非常大的内窥镜图像数据集(=500.000)来预训练卷积神经网络的好处。随后，使用几个特定于 BE 的数据集来微调模型，每次都更接近目标域。他们将 CAD 结果与四个不同技能水平的 53 名医生进行了比较，结果表明 AI 算法的表现优于每位医生[50]。在后来的研究中，相同的算法在临幊上进行了现场评估，结果非常好，其中 90% 的患者通过该算法正确分类[52, 53]。同一小组还研究了 AI 在内窥镜视频方面的挑战和机遇[54]。在 EAC 的 WLE 图像上训练的 CNN 的任务也是在内窥镜视频中发现病变。虽然结果很有希望，但非信息框架被证明是一个问题。第二个问题是，当只有一小部分实际上是癌变时，如何确定整个视频的诊断，误报使这成为一个难题。最后，对所谓的“场效应”进行了一项研究，该效应认为，发育不良 BE 中的非异常增生组织表现出类似的特征，这些特征不像实际病变那样明显，但可能被 AI 感知。在 Verhage 等人[55]的一项研究中，CNN 用于发育不良图像的非发育不良斑块，并与非发育不良的巴雷特食管斑块进行比较。结果显示两者之间存在统计差异，这表明可能存在视野效应。

表 1 相关参考文献概述 - 其中，各缩写的含义分别是 C: 分类，S: 分割，D: 检测，R: 重建，CNN: 卷积神经网络，SSD: 单次检测器，EOCT: 内窥镜光学相干断层扫描，CAM: 类激活图，RNN: 递归神经网络，CV: 交叉验证，LOO: 省略一个，LOPO 遗漏一名患者，Tr: 培训，Val: 验证，Te: 测试，PP-CV: 每个患者的交叉验证，WLE: 白光内窥镜检查，NBI: 窄带

成像, AAC: 醋酸色内窥镜, pCLE: 基于探针的共聚焦激光内窥镜, VLE: 体积激光显微内窥镜

参考文献					
WLE	任务	描述	验证类型	模态	数据集
Vander Som men 2016[25]	C+S	Gabor features and SVM	LOO-CV	WLE	End oVis
Souza 2018[26]		Color co-occurrence and SVM	LOPO-CV	WLE+NB I	End oVis+Aus burg(in house)
Riel 2018[31]	C	CNN codes	LOPO-CV	WLE	End oVis
Ghat wary 2019[27]		CNN+Gabor	LOPO-CV	WLE	End oVis+Kvasir
Mendel 2017[36]		Fine-tune Resnet	LOPO-CV	WLE+NB I	End oVis
Eb igbo 2020[38]	C+S	Resnet+Deep lab V 3+	Clinical	WLE	-
Souza 2020[42]	C	GAN data augmentation	Tr/Val	WLE	End oVis+Aus burg(in house)
Ghat wary 2019-III[43]	C+D	R-CNN and SSD	LOPO-CV	WLE	End oVis
Ghat wary 2020[481]	C+D	R-CNN and SSD	Tr/Val/Te	WLE	End oVis
Hashimoto 2020[49]	C+D	Xception+Yolo	Tr/Val	WLE+NB I	In-house dataset
Vander Putten 2019[51]	C+S	Domain specific pre training and CNN	Tr/Val/Te	WLE	In-house dataset
DeG roof 2020[50]	C+S	U-Net/Resnet hybrid	Tr/Val/Te	WLE	In-house dataset
Vander Putten 2020[52]	C+S	U-Net/Resnet hybrid	Clinical	WLE	-
DeG roof 2020-III[53]	C+S	Clinical evaluation ofdeveloped model	Clinical	WLE	-
Vander Putten 2020-II[54]		Endoscopic videoclassification	PP-CV	WLE	In-house dataset
Verh age 2020[55]		Field effectclassification	PP-CV	WLE	In-house dataset
NBI	任务	描述	验证类型	模态	数据集

Rajan 2009[63]		Multimodality classification with SVM/KNN	CVWLE+NB I+AAC		In-house dataset
Vander Putten 2019-II[64]	C	NBI-zoom video characterization using Resnet model	PP-CV	WLE+NB I	In-house dataset
Struyvenberg 2020[66]	C	NBI-zoom video characterization using Resnet model	PP-CV	NBI	In-house dataset
Endo microscopy	任务	描述	验证类型	模态	数据集
Grisan 2012[68]	C	Two-stage classification with feature extraction	LOO-CV	pCLE	In-house dataset
Veronese 2013[69]	C	Hybrid image and patch-based feature extraction	LOO-CV	pCLE	In-house dataset
Hong 2017[70]		CNN for classification of endo microscopy images	Tr/Val	pCLE	IS BI 2016challenge
Ghatwary 2017[71]	C	Feature extraction+SVM for endo microscopy images	LOO-CV	pCLE	In-house dataset
Ghatwary 2019- III[72]	C	Feature enhancement+ SVM	LOO-CV	pCLE	In-house dataset
Pulido 2020[73]		BE screening using endo microscopy videos	Tr/Val	pCLE	In-house dataset
Qi2004[75]	C	E OCT classification using CSAC	LOO-CV	E OCT	In-house dataset
Qi2006[76]	C	E OCT classification using Fourier analysis	LOO-CV	E OCT	In-house dataset
Rodriguez 2015[77]	C	GL CM+bayes classifier	LOO-CV	VLE	In-house dataset
Swager 2017[96]	C	Clinically inspired classification of VLE	CV	VLE	In-house dataset

Scheeve 2019[78]	C	Clinically inspiredgl and s feature classification	Tr/Val/Te	VLE	In-house dataset
Vander Som men 2018[79]	C	Multiple handcrafted features+extensive classifier analysis	LOO-CV	VLE	In-house dataset
Putten 2019-III[80]	S	Region of interest segmentation with U-Net	Fivefold CV	VLE	In-house dataset
Fonal a 2019[81]	C	Ensemble of CNN s+CAM	Tr/Val/Te	VLE	In-house dataset
Putten 2020- IIII[82]	C	Principal dimensioned coding class if cation	Fivefold CV	VLE	In-house dataset
Quality assessment	任务	描述	验证类型	模态	数据集
Hwang 2005[84]		GL CM+SURF bubble classification	Twofold CV	WLE	In-house dataset
Pie tr 2018[86]	C	DFT+frame clustering	-	WLE	In-house dataset
Wang 2019[85]	C	RSS filter	Tr	WLE	In-house dataset
Akbari 2018[87]	C+S	HSV+RGB features	Tr	WLE	CVC-Colon DB
Van Dongen 2016[88]	C	DCT+color+blur-based features	Tenfold CV	WLE	In-house dataset
Taj bakhsh 2016[89]	C	Full training versus fine-tuning	Tr/Val	WLE	In-house dataset
Islam 2018[90]	C	Bubble, water, and blurry frame detection+CNN	Tr/Val	WLE	In-house dataset
Hong 2014[91]	R	3D colon reconstruction with depth from intensity	-	WLE	In-house dataset
Putten 2019- IVV[92]	C	Informative frame classification with CNN+HMM	Fivefold CV	WLE	In-house dataset

Boers 2020[93]	C	Three informative frame classification with RNN	Fivefold CV	WLE	In-house dataset
Boers 2020- I[94]	C	Tissue classification with RNN	Fivefold CV	WLE	In-house dataset