

计算机辅助诊断系统在甲状腺结节 超声诊断中的应用进展

王琳萍 张 波

摘 要 甲状腺结节是常见的临床问题。超声检查是其首选的影像学诊断工具,用于评估甲状腺结节患者的恶性风险,结合结节大小、局部侵犯情况、颈部淋巴结是否转移等推荐细针穿刺活检(fine-needle aspiration, FNA),并且在患者治疗后继续管理方案的决策中发挥不可或缺的作用。超声的正确高效使用,不仅可以减少不必要的FNA和(或)诊断性手术,还能减轻患者的焦虑及医疗保健系统的负担。近年来,以机器学习和深度学习为核心,基于人工智能的甲状腺结节计算机辅助诊断系统已取得巨大进步,在一定程度上克服了超声诊断中的操作者依赖性,但是受到数据源不同、算法不同等因素的影响,各中心的结果在跨中心运用时表现并不稳定。本文对其应用现状及研究进展进行综述。

关键词 人工智能 计算机辅助诊断系统 甲状腺结节 超声诊断

中图分类号 R581

文献标识码 A

DOI 10.11969/j.issn.1673-548X.2021.10.035

甲状腺结节是常见的内分泌肿瘤,在无症状人群中约占19%~67%,但大部分结节为良性,恶性结节不足10%^[1]。如何将甲状腺癌从高发的甲状腺结节中甄别出来,就显得至关重要。

超声是甲状腺结节首选的影像学诊断方法,所有已知或可疑甲状腺结节的患者均应行超声检查^[2]。超声用于评估恶性肿瘤的风险,结合结节大小、局部侵犯情况、颈部淋巴结是否转移等推荐细针穿刺活检,在甲状腺结节的全程管理中发挥重要作用^[3]。然而,对于经验不足的检查者来说,准确识别和解读超声特征具有挑战性,易引起观察者间和观察者内部的差异,导致不必要的细针穿刺和(或)诊断性手术,故准确诊断甲状腺结节良、恶性既能让患者得到及时治疗,也能减少医疗资源的浪费和缓解患者不必要的紧张、焦虑情绪。

一、甲状腺成像中基于人工智能的计算机辅助诊断系统的开发

为减少甲状腺结节超声诊断的人员依赖性,规范化超声图像的解读,自21世纪初以来,风险评估系统(risk-stratification systems, RSS)开始用于甲状腺结

节超声诊断的有效管理,并由最初简单分类的定性分级系统发展为定量评分系统,即甲状腺影像报告与数据系统(thyroid imaging reporting and data system, TI-RADS),以可疑超声指标的数量和指标风险评分为依据,进行甲状腺结节的恶性风险分层^[4]。

当今各种TI-RADS分类系统已在实践中得到广泛应用。但不足之处在于具有不同风险的IV类和V类甲状腺结节的临床处置相同,所以研究者建议采取分级更精细的RSS,以便对甲状腺结节进行个性化和优化管理^[5]。为满足这一临床需求,以机器学习(machine learning, ML)和深度学习(deep learning, DL)为核心,基于人工智能(artificial intelligence, AI)的甲状腺结节计算机辅助诊断系统被引入临床。AI可以提取和量化关键的图像信息,从而使图像诊断从主观的定性分析转化为客观的定量分析,因此基于AI的CAD系统可提高医生超声诊断准确性并解决分级RSS复杂性的问题,以避免低危结节进行不必要的FNA^[6]。

二、计算机辅助诊断系统在甲状腺结节超声诊断中的应用现状及研究进展

1. 机器学习及深度学习:基于AI的CAD系统具有两种核心技术:ML和DL。使用ML技术时,通常是基于超声特征(例如成分、形状、边缘、回声和钙化)开发CAD系统,并证明其诊断甲状腺癌的潜力。基于ML的CAD系统可以为医生提供参考意见。与ML比较,DL则不需要医生预处理图像及选择特征,

基金项目:北京市自然科学基金资助项目(7192152);北京协和医学院学科建设项目(清华211)(201920102305);北京协和医学院研究生教育教学改革项目(2018zlgc0108)

作者单位:100029 北京,中日友好医院(王琳萍、张波);100730 中国医学科学院/北京协和医学院(王琳萍)

通讯作者:张波,教授,博士生导师,电子邮箱:zora19702006@163.com

降低了对数据质量的要求,可更客观地分析图像信息。近期使用 DL 进行的研究都开发了分类模型,但未提供有关超声特征的信息^[7]。DL 的算法目前还属于黑盒子状态,其对甲状腺结节检出原理尚无法解释。

2. 甲状腺结节超声诊断的商业化 CAD 系统:目前,针对软件植入和外部验证问题,已开发出两种用于甲状腺结节良恶性诊断的商业化 CAD 系统。AmCAD-UT(中国台湾地区 AmCAD Biomed 公司)是用于超声诊断甲状腺结节的商业化 CAD 系统,旨在使用统计模式识别和量化算法来表征甲状腺结节,并根据 TI-RADS 分类提供恶性肿瘤的风险。用于甲状腺的 S-Detect(韩国 Samsung Medison 公司)是另一种商业化的 CAD 系统,应用最为广泛,已集成到商用超声平台中。它使用二分类结果(可能是良性或恶性)或 TI-RADS 分类结果来显示超声特征和可能的诊断。甲状腺的 S-Detect1 利用基于支持向量机模型的 ML 技术,甲状腺的 S-Detect2 利用基于卷积神经网络的 DL 技术。S-Detect 技术诊断甲状腺的准确性较高,该技术有助于提高低年资医生诊断的特异性和准确性,与高年资医生诊断一致性相对较好,未来有助于超声图像的标准化判读^[8,9]。

3. CAD 系统在甲状腺结节超声诊断中的应用现状:随着技术的进步,CAD 系统在甲状腺结节超声诊断中的应用,从只可重点标注恶性程度较高的特征,到实现对整个甲状腺结节超声图像的分析及自动诊断。DL 出现之后,计算机辅助甲状腺结节超声诊断系统在临床上的应用取得了质的飞跃。

单一超声征象计算机化:第一阶段的应用主要局限在对甲状腺结节单个特征的自动化识别方面,从而实现良恶性判断。单一超声征象计算机软件可自动计算感兴趣区域内的某个特征值(如回声、钙化等),并输出量化参数值,从而预测甲状腺结节良恶性。Choi 团队的研究中,超声医生使用计算机软件手动选择甲状腺结节区域,软件自动识别钙化并计算一些钙化指数,从而得出量化钙化更客观、更少依赖于操作者,可以提高检测的敏感度的结论^[10]。Meta 分析发现最广泛使用的特征是形状、边缘、回声、钙化、成分和大小,这些特征应当成为 AI 模型学习的重点。纵横比 > 1 表明甲状腺结节在空间和方向生长上的变异,是恶性肿瘤最具提示性的特征^[11]。

自动检测及自动诊断:随着 ML 技术的发展,CAD 系统研究已经能实现对整个甲状腺结节超声图

像的分析,而不仅限于对单独超声特征的定量分析。该技术通常包括图像预处理、特征提取和数据分类。其中图像分类即分类器通过数据输入、监督学习、训练和反馈等建立未知图像的自动分类模型。最常见的两种分类器是支持向量机(support vectors machine, SVM)和人工神经网络(artificial neural network, ANN)。

SVM 旨在间隔最大化的前提下找到一条最优的分割线进行二分类。基于 SVM 的图像分析系统用于评估甲状腺结节的恶性风险时,可从每个结节中自动计算出 40 个纹理特征,并与 SVM 算法一起用于图像分析系统的设计,最高分类准确率为 98.65%^[12]。尤其是在甲状腺结节数据集信息较少的情况下,SVM 对甲状腺结节的检出对计算机硬件要求较低,成本也低于深度学习;但采用人为设计的特征提取进行选择,易造成部分结节信息丢失。深度学习则可同时完成结节定位、分割和分类。

卷积神经网络(convolutional neural network, CNN)属于深度学习的一种,无需手动选取特征,训练好权重即得特征分类。Kwon 等^[13]采用迁移学习方法,用 762 例患者的甲状腺横、纵向超声图像预先训练深度学习模型 VGG16,从而使用 CNN 为甲状腺超声图像分类构建高效、准确的 CAD 系统。在定位上,由于 CNN 具有图像特征位移不变性,学习到的特征可从图像的不同位置中提取出来,不会因结节位置多变和体积较小导致检出率下降,具有较好的泛化能力。在结节分类中,深度学习实现从原始图像输入到最终分类的映射,消除 CAD 手工设计特征对最终分类的影响。虽然 CNN 对图像物体的分类与定位具有较强的优势,但对甲状腺结节的检出也存在不足,需对医生标注过的甲状腺图像进行学习,而医生标注质量的高低则直接决定了学习质量及模型的效能。

4. CAD 系统在甲状腺结节超声诊断中的研究进展:基于 AI 的 CAD 系统与 RSS 在商用超声机器上的集成,可减少图像判读过程中操作人员的依赖性,有助于实时解读以评估甲状腺结节患者的恶性肿瘤风险和是否进行 FNA,但 CAD 系统的实际临床意义需要在不同的临床环境中进一步验证。

Choi 等^[14]应用 S-Detect™ 模式,纳入 89 例患者共 102 个甲状腺结节,研究发现经验丰富的诊断医生对甲状腺癌的诊断特异性及特征曲线下面积明显高于 CAD 系统,但两种方法的诊断敏感度比较差异无统计学意义。另一项研究选取 50 例患者共 117 个甲

状腺结节进行诊断, CAD 系统的敏感度和特异性与诊断医生比较差异无统计学意义, 提示当 CAD 系统用于辅助诊断医生时, 医生诊断敏感度提高, 特异性下降^[15]。Gao 等^[16]利用基于多尺度 CNN 模型建立的甲状腺癌 CAD 系统, 诊断甲状腺癌的敏感度、特异性、准确度分别为 96.7%、48.5% 和 82.2%。具有 20 年甲状腺超声经验的一位医生利用 Kwak - TIRADS、美国甲状腺协会风险分层、美国放射协会 - TIRADS 这 3 种指南单独对所有结节进行诊断比较, 发现 CAD 系统的敏感度接近超声医生, 而特异性稍低。Gitto 等^[17]研究发现, CAD 系统与有经验的超声医生在方向、成分、回声方面和 Kwak - TIRADS 的观察者一致性好, 而在边缘方面的一致性差。该研究还比较了 CAD 系统和医生的诊断能力, 结果发现诊断医生对需要随访或手术的甲状腺结节的诊断敏感度明显高于 CAD 系统, 特异性差异无统计学意义。这与 Choi 等的研究结论大相径庭, 或许是由于两个研究使用的参考标准不同, 且该研究中进行 CAD 系统处理的两位医生诊断甲状腺结节经验不足。Li 等^[18]开发了含超过 30 万张图像的分类模型, 直接比较用 TI - RADS 诊断的医生和使用二分类结果的 CAD 系统, 结果提示, 新开发的 CAD 系统与熟练的医生具有相似的敏感度 (分别为 84.3% ~ 93.4% vs 89.0% ~ 96.9%) 以及更高的特异性 (分别为 86.1% ~ 87.8% vs 57.1% ~ 68.6%)。这一结论存在争议, 例如与此前的研究报道比较, 诊断医生的特异性相对较低 (57.1% ~ 68.6%), 以及使用 TI - RADS 分类的诊断医生与使用二分类结果的 CAD 系统之间是否可直接比较有待商榷。但是这项研究在技术性方面的成功值得引起研究者的注意, 该研究亟待进一步在不同的地理环境中进行验证。

后续的研究着眼于医生的经验是否会影响 CAD 系统的诊断结果。Chung 等^[19]探讨 CAD 系统在甲状腺结节诊断中是否优于 3 位具有不同经验 (1 个月、4 年和 7 年甲状腺超声经验) 的诊断医生。CAD 系统的诊断准确率 (88.5%, 95% CI: 82.7 ~ 92.5) 不低于甲状腺超声经验较少 (1 个月和 4 年) 的诊断医生 (83.0%, 95% CI: 76.5 ~ 88.0, $P = 0.000$), 而低于有经验的诊断医生 (7 年) (95.8%, 95% CI: 91.4 ~ 98.0, $P = 0.138$), 该系统可为经验较少的甲状腺超声工作者提供甲状腺恶性结节诊断决策支持。一项多中心前瞻性研究中, 494 例患者 (565 个甲状腺结节) 于 2019 年 1 ~ 9 月在 4 家医院接受了超声检查后

的手术或活检。计算 CAD 系统和诊断医生的诊断准确率指标, 并与病理结果进行比较。结果表明 CAD 系统的诊断准确率与高年资诊断医生相当, 特异性高于低年资诊断医生 (87.5% vs 70.4%, $P = 0.03$)。CAD 系统的敏感度低于高年资和低年资诊断医生, 但差异无统计学意义 (76.9% vs 86.9%, $P > 0.5$; 76.9% vs 82.6%, $P > 0.5$)^[20]。

综上所述, 不同版本的 CAD 系统诊断效率接近或略低于高年资医生, 但是明显高于低年资医生。然而, 由于分化型甲状腺癌占甲状腺癌总体比例 90% 以上, 预后好, 病死率低, 许多研究者倾向于将高特异性作为减少不必要的 FNA 的方法。因此可以使用 CAD 系统作为高敏感度的筛查工具, 以协助初级医疗中心经验较少的操作员。关于是否进行 FNA 可参考甲状腺成像专家的建议从而增加诊断的特异性。总之, 计算机辅助诊断系统对低年资诊断医生具有重要的辅助诊断价值, 其在临床的应用具有广阔的前景, 然而它的应用也面临很多挑战及局限性。

三、挑战及局限性

CAD 技术的局限性和超声检查的性质对超声 AI 的发展和临床应用提出了挑战。(1) CAD 技术无法全面评估血流和硬度等信息, 也无法将症状和体征纳入考虑范围内。(2) AI 的输出结果通常以概率形式呈现, 概率尺度的准确性很大程度上受到基线特征和临床环境的影响。用于医学诊断和预测的 AI 算法的可推广性有限, 不同患者和医院之间的 AI 准确性具有显著差异, 即“过拟合”。这种局限性在超声检查的 AI 算法上尤其明显, 因为超声检查广泛使用于各种临床环境和患者, 由具有不同专业知识的各种医疗专业人员执行, 超声的操作者依赖性引入了额外变异。超声检查系统也比 CT 或 MRI 更加多样化, 有更多的供应商和版本。故扫描和图像采集的标准化对于 AI 在超声中的成功应用至关重要。(3) 即使系统是使用相对较小的样本量开发的, 但在大多数情况下仍缺乏测试和验证数据集。近期一项回顾医学影像诊断 AI 算法的系统评价分析发现, 只有 6% 发表在同行评议期刊上的此类研究进行了某种形式的外部研究验证 (无论在方法上是否充足)。(4) 虽然人们期望 AI 有助于经验不足的检查者进行诊断, 然而未接受专业培训的人员在使用 AI 时会出现解读困难, 且倾向于未进行必要的评估便直接采纳 AI 的结果, 这最终将影响超声检查报告的准确性。虽然将

AI 引入医学的路上仍然困难重重,但是挑战就是机遇,突破就会成长。

四、展 望

医疗领域的 AI 研究正在迅速发展并出现了许多潜在的应用前景。首先,超声对操作者的依赖性使得验证 AI 的前瞻性研究显得尤为必要。AI 的作用不仅取决于技术分析能力,还取决于计算机化结果如何呈现给医师进行诊断。考虑到在采集超声图像以及实时检查过程中结合 AI 做出决策时的操作人员依赖性,回顾性收集图像的分析与真实临床环境之间比较差异可能有统计学意义。到目前为止,超声 AI 的研究主要集中在回顾性研究,应该进行更多涉及检查者和 AI 系统之间的实际交互的前瞻性研究。其次,过度拟合的问题强调了 AI 算法在各种实际临床环境中的充分外部验证的重要性。未来超声的 AI 研究除了开发新的算法外,还应强调对已开发算法的外部验证。严格的外部验证有助于明确 AI 算法何时可保持其预期精度,从而有助于确保医生安全有效地使用 CAD 系统^[21]。

综上所述,对于诊断医生来说,了解 CAD 技术的优缺点,最大限度地利用其辅助诊断甲状腺病变的功能是非常重要的。基于对超声检查的性质的认识,未来需要关注以下问题。超声检查本身和相关的临床和流行病学方面的知识、扫描和图像采集的标准化、AI 算法充分外部验证等都尤为重要。还应进行涉及操作者和 AI 系统之间实际相互作用的前瞻性研究,而不仅仅是分析回顾性收集的图像。AI 初入临床实践便已取得可观的成果,应用于超声检查的具体例子在未来还会继续增加,相信计算机辅助诊断系统终将成为诊断医生的第三只眼。

参考文献

- 1 Durante C, Grani G, Lamartina L, *et al.* The diagnosis and management of thyroid nodules: a review[J]. *JAMA*, 2018, 319(9): 914-924
- 2 Haugen BR, Alexander EK, Bible KC, *et al.* 2015 American Thyroid Association Management Guidelines for adult patients with thyroid nodules and differentiated thyroid cancer; the American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid cancer[J]. *Thyroid*, 2016, 26(1): 1-133
- 3 韩志江, 吴颖为. 甲状腺结节影像检查流程专家共识[J]. *中华放射学杂志*, 2016, 50(12): 911-915
- 4 Jeong EY, Kim HL, Ha EJ, *et al.* Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators[J]. *Eur Radiol*, 2019, 29(4): 1978-1985
- 5 Ting DSW, Cheung CY, Lim G, *et al.* Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes[J]. *JAMA*, 2017, 318(22): 2211-2223
- 6 Ha EJ, Baek JH, Na DG. Risk stratification of thyroid nodules on ultrasonography: current status and perspectives[J]. *Thyroid*, 2017, 27(12): 1463-1468
- 7 Zhang B, Tian J, Pei S, *et al.* Machine learning-assisted system for thyroid nodule diagnosis[J]. *Thyroid*, 2019, 29(6): 858-867
- 8 陈晨, 孙鹏飞, 郭君, 等. AI 在甲状腺结节良恶性中的诊断价值[J]. *中国超声医学杂志*, 2020, 36(7): 585-588
- 9 李潜, 刘春丽, 韦雅楠, 等. S-Detect 技术在甲状腺可疑结节诊断中的应用价值[J]. *中华超声影像学杂志*, 2020, 29(11): 964-968
- 10 Choi WJ, Park JS, Kim KG, *et al.* Computerized analysis of calcification of thyroid nodules as visualized by ultrasonography[J]. *Eur J Radiol*, 2015, 84(10): 1949-53
- 11 Xia J, Chen H, Li Q, *et al.* Ultrasound-based differentiation of malignant and benign thyroid Nodules: an extreme learning machine approach[J]. *Comput Methods Programs Biomed*, 2017, 147(2017): 37-49
- 12 Shankar K, Lakshmanaprabu SK, Gupta D, *et al.* Optimal feature-based multi-kernel SVM approach for thyroid disease classification[J]. *J Supercomput*, 2018, 76(2): 1128-1143
- 13 Kwon SW, Choi IJ, Kang JY, *et al.* Ultrasonographic thyroid nodule classification using a deep convolutional neural network with surgical pathology[J]. *J Digit Imaging*, 2020, 33(5): 1202-1208
- 14 Choi YJ, Baek JH, Park HS, *et al.* A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: initial clinical assessment[J]. *Thyroid*, 2017, 27(4): 546-552
- 15 Yoo YJ, Ha EJ, Cho YJ, *et al.* Computer-aided diagnosis of thyroid nodules via ultrasonography: initial clinical experience[J]. *Kor Jou Radiol*, 2018, 19(4): 665-672
- 16 Gao L, Liu R, Jiang Y, *et al.* Computer-aided system for diagnosing thyroid nodules on ultrasound: a comparison with radiologist-based clinical assessments[J]. *Head Neck*, 2018, 40(4): 778-783
- 17 Gitto S, Grassi G, De Angelis C, *et al.* A computer-aided diagnosis system for the assessment and characterization of low- to high suspicion thyroid nodules on ultrasound[J]. *Radiol Med*, 2019, 124(2): 118-125
- 18 Li X, Zhang S, Zhang Q, *et al.* Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study[J]. *Lancet Oncol*, 2019, 20(2): 193-201
- 19 Chung SR, Baek JH, Min KL, *et al.* Computer-aided diagnosis system for the evaluation of thyroid nodules on ultrasonography: prospective non-inferiority study according to the experience level of radiologists[J]. *Korean J Radiol*, 2020, 21(3): 369-376
- 20 Ye FY, Lyu GR, Li SQ, *et al.* Diagnostic performance of ultrasound computer-aided diagnosis software compared with that of radiologists with different levels of expertise for thyroid malignancy: a multicenter prospective study[J]. *Ultrasound Med Biol*, 2021, 47(1): 114-124
- 21 Ha EJ, Baek JH. Applications of machine learning and deep learning to thyroid imaging: where do we stand? [J]. *Ultrasonography*, 2021, 40(1): 23-29

(收稿日期: 2021-04-11)

(修回日期: 2021-04-23)