



· 综 述 ·

乳腺筛查网络测评平台 (BREAST) 在提高放射科医师诊断能力中的应用进展

孙诗韵^{1,2}, 肖 勤¹, Tong Li³, 王思敏¹, 黄 琰¹, 顾雅佳¹

1. 复旦大学附属肿瘤医院放射诊断科, 复旦大学上海医学院肿瘤学系, 上海 200032 ;

2. 上海市影像医学研究所, 上海 200032 ;

3. 悉尼大学医学与健康学院公共卫生学院, 澳大利亚 悉尼 2066

[摘要] 乳腺筛查网络测评平台 (BreastScreen Reader Assessment Strategy, BREAST) 是一种乳腺影像研究和培训工具, 通过提供大量典型的病例集, 并借助在线应用程序Breast Score进行实时、个性化分析和反馈, 从而帮助放射科医师提高诊断能力和水平。目前, 该平台已在多个国家中得到了应用和发展, 并且均取得了不错的效果。因此, 本文简要介绍BREAST, 以及BREAST在乳腺癌筛查中的应用情况及其优劣势, 以帮助放射科医师了解和应用BREAST平台。

[关键词] 乳腺癌; 乳腺筛查网络测评平台; 放射科医师; 诊断能力

DOI: 10.19732/j.cnki.2096-6210.2022.04.016

中图分类号: R737.9; R445.4 文献标志码: A 文章编号: 2096-6210(2022)04-0444-05

The application progress of BreastScreen Reader Assessment Strategy (BREAST) in improving the diagnostic ability of radiologists

SUN Shiyun^{1,2}, XIAO Qin¹, TONG Li³, WANG Simin¹, HUANG Yan¹, GU Yajia¹

(1. Department of Radiology, Fudan University Shanghai Cancer Center, Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China; 2. Shanghai Institute of Medical Imaging, Shanghai 200032, China; 3. School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney 2066, Australia)

Correspondence to: GU Yajia E-mail: cjr.guyajia@vip.163.com

[Abstract] The BreastScreen Reader Assessment Strategy (BREAST) is a breast imaging research and training tool that provides real-time, personalized analysis and feedback with a large, typical patient set through the online application Breast Score. Help radiologists improve their diagnostic capabilities and skills. At present, the platform has been applied and developed in many countries, and all have achieved good results. Therefore, this review briefly introduced the breast platform and its application in breast cancer screening, as well as its advantages and disadvantages, so as to help radiologists understand and apply the breast platform.

[Key words] Breast cancer; BreastScreen Reader Assessment Strategy; Radiologist; Diagnostic ability

作为威胁全球女性生命健康的首要恶性肿瘤, 乳腺癌的发病率逐年增高^[1]。近年来, 在发达国家中, 随着乳腺癌筛查项目的实施和普及, 其死亡率呈现出逐年下降的趋势^[1-2]。多项研究^[3-4]已证实, 乳腺X线摄影检查作为乳腺癌早期筛查的重要手段之一, 可有效地提高乳腺癌的检出率, 进而降低死亡率。乳腺癌的早期发现很大程度上取决于放射科医师对于乳腺X线摄影

图像的准确判断。因此, 着力提高放射科医师的诊断能力显得尤为重要。

乳腺筛查网络测评平台 (BreastScreen Reader Assessment Strategy, BREAST) 是一种乳腺影像研究和培训工具^[5], 旨在通过实时在线评估和反馈, 提高放射科医师的诊断能力和水平。本文主要对该平台在提高放射科医师诊断能力中的应用进行综述。

基金项目: 中华国际医学交流基金会2020sky影像科研基金 (Z-2014-07-2003-06) ; 国家自然科学基金 (82071878) ; 促进市级医院临床技能与临床创新三年行动计划-重大临床研究项目 (SHDC2020CR2008A) ; 上海市抗癌协会“翱翔”计划 (SACA-AX-201903)

通信作者: 顾雅佳 E-mail: cjr.guyajia@vip.163.com

1 影响乳腺癌筛查准确度的因素

乳腺X线摄影检查已被证实是乳腺癌筛查的有效工具,可以有效地提高乳腺癌的检出率,降低死亡率^[3-4]。然而,研究^[6-7]显示,在筛查时5%~30%的乳腺癌会被遗漏(即假阴性),而诊断错误率可能高达50%。Brem等^[8]及Gandomkar等^[9]的研究显示,导致诊断错误的原因60%~63%来源于认知偏差,即对于乳腺X线摄影的正常表现认识不足,又无法准确识别其异常表现,从而导致了假阴性、假阳性或过度诊断。假阴性会延误癌症的早期发现和治疗,假阳性则会引起患者焦虑并导致额外检查费用的产生,过度诊断可能导致过度治疗。因此,提高乳腺筛查的准确度显得尤为重要。

然而,不同国家的放射诊断科医师对于乳腺X线摄影的筛查准确度各不相同,发展中国家普遍低于50%而发达国家在80%以上^[10-11]。研究^[12-14]显示,差异主要来源于以下3种原因:①患者群体差异,例如不同地区的患者乳腺密度不同。致密的乳腺组织很可能掩盖微钙化、结构扭曲等有助于诊断的影像学表现,从而导致乳腺癌的检出率和诊断灵敏度下降^[15-16];②诊断医师差异,例如工作年限、乳腺X线摄影片的阅片年数、每周的阅片时间、阅片量、是否接受过乳腺X线摄影培训、是否参与乳腺筛查工作等;③技术和环境差异,例如乳腺X线摄影片的质量、诊断环境、阅片设备等。其中,来自诊断医师的差异是影响筛查准确度的主要因素^[17-18]。研究^[19-20]显示,放射科医师的诊断能力主要与工作年限、阅片年数、阅片量,以及是否参与乳腺筛查工作、接受过乳腺X线摄影培训有关。例如,一项回顾性研究^[21]显示,在初次筛查时一些乳腺癌患者就已经表现出某些细微并可疑的影像学表现,但由于这些影像学表现不典型,很容易被经验不足的医师所忽视,从而导致了假阴性诊断,耽误了疾病的早期发现和治疗。

为了减少乳腺筛查过程中假阳性和假阴性产生,许多国家的乳腺癌筛查质量保证计划要求定期进行临床审核和反馈^[22-23],但这种反馈往往需要2年的临床数据才能提供较为准确的信息。

因此,近年来越来越多的国家鼓励放射诊断科医师进行继续医学教育来评估并提高自身的诊断能力^[23]。目前,澳大利亚和英国已经成功搭建了乳腺X线摄影的教育培训平台(BREAST^[24]和PERsonal performance^[25]),这些平台通过组建测试集来进行技能训练和评估,并可及时提供分析和反馈,帮助测试者提升诊断能力。

2 BREAST平台介绍

2011年,悉尼大学的一项研究^[26]表明,在乳腺癌筛查中约44%的乳腺病变会被遗漏,并且诊断医师的灵敏度普遍低于70%^[27]。有鉴于此,悉尼大学和新南威尔士癌症研究所的资深放射科诊断医师们开发并建立了BREAST^[5]。该平台拥有丰富的病例测试集,并借助在线应用程序Breast Score为放射诊断科医师提供实时、个性化的分析和反馈,帮助放射诊断科医师提高诊断能力。在过去7年里,BREAST一直被当作澳大利亚乳腺癌筛查项目的官方质量培训工具,并吸引了超过80%的医师参与,被誉为提高乳腺癌筛查能力的最佳工具^[5]。

2011—2017年,该平台共发布了7套乳腺X线摄影测试集,每套测试集包含了60个案例(40例正常乳腺和20例乳腺癌)。每套测试集的X线图像均来自于澳大利亚乳腺癌筛查图像数据库,包括双侧头尾位和内外侧斜位片。所有阳性病例均由活检证实,阴性病例由2名放射科医师报告并随访2年仍为阴性来确定。BREAST平台尽可能平衡了乳腺密度、病灶类型、位置、大小等因素对诊断带来的影响。在所有测试集中,高密度乳腺和低密度乳腺分别占55%和45%;大病灶(>10 mm)和小病灶(<10 mm)分别占54%和49%;病灶位于左侧者占51%,右侧者占52%。病灶类型包括毛刺状肿块(39%)、钙化(20%)、多发肿块(13%)、不对称(17%)、结构扭曲(7%)以及混合型(11%)。

在测试之前,所有参与的放射科医师都需要接受一项关于年龄、工作时间、工作经验等个人信息的简短调查。在测试过程中,受试者可以自由使用平移、缩放等后处理工具,并根据乳腺

影像报告和数据系统 (Breast Imaging Reporting and Data System, BI-RADS) 对可疑病灶进行分类和勾画。测试结束后, 系统会立即显示每例患者阳性病灶的位置、数量以及适合的BI-RADS分类, 并根据受试者的表现提供包括灵敏度、特异度、受试者工作特征 (receiver operating characteristic, ROC) 曲线的曲线下面积等评估指标。

3 BREAST平台的应用情况

从2011年成立至今, BREAST平台已在10余个发达国家和发展中国家得到了应用和推广, 范围涵盖南亚、东南亚、大洋洲、北美洲和欧洲, 并与哈佛医学院、复旦大学、越南国家癌症中心、新加坡国立大学、约旦科技大学、德黑兰医科大学等10余个国际研究机构建立了合作伙伴关系^[20, 23, 28-30]。多项研究^[31]显示, 由BREAST提供的乳腺X线摄影测试集与临床审核之间拥有良好的相关性, 并且定期参加BREAST测试可以使放射科医师的诊断灵敏度和准确度提高6%~34%^[23-24, 29, 32], 这表明BREAST是评估和提高放射科医师诊断能力的有效工具。

在发达国家中, 澳大利亚是最先开发并使用BREAST的国家。Suleiman等^[24]的研究显示, 14名澳大利亚放射科医师在连续3年参与了BREAST测试后, 诊断能力平均提高了34%, 并且达到了继续教育的目的。Trieu等^[23]的研究进一步证实了这一点。50名放射科医师中约83%的医师表示在第1次完成BREAST测试后诊断能力得到了提升, 并且随着测试次数的增加, 诊断病灶的灵敏度也随之提高。与第1次测试相比, 第2、3次测试的诊断灵敏度增加了20%~31%。第3次测试结束后, 几乎所有人都获得了更高的JAFROC分数。这表明由实践和反馈系统组成的BREAST测试集可以有效地提高澳大利亚放射科医师的乳腺癌筛查能力。

Li等^[19]比较了澳大利亚和中国香港及广东省地区的年轻放射科医师在诊断致密型乳腺患者中的诊断表现, 结果显示澳大利亚医师拥有更高的诊断灵敏度和JAFROC分数 (分别高出39%和40%), 并且在病灶定位方面拥有更出色的表

现。此外, 许多研究^[33]指出放射科医师的诊断能力会随着工作经验和年限的增加而有所提高。但Li等^[19]的研究显示工作经验和诊断灵敏度无关 ($P>0.05$), 而参与教育培训后则可明显提高诊断灵敏度 ($P<0.05$)。Li等^[19]认为澳大利亚医师接受了专门的乳腺X线摄影诊断培训计划 (如BREAST), 并且拥有较多乳腺癌筛查的经验, 因此更擅长于发现无症状女性中的微小肿瘤^[14]。相比之下, 在中国仍然没有乳腺癌筛查计划和专业的乳腺X线摄影教育培训平台, 缺乏完善的学习和反馈机制, 因而诊断能力也相对较低。这与亚洲其他发展中国家的现状相似, Jackson等^[10]的研究显示, 越南放射科医师的乳腺癌筛查率远低于澳大利亚和新加坡 (47% vs 74%和80%)。更重要的是, 研究^[11]发现, 越南医师对于毛刺或星芒状肿块的检出率低于钙化、多发肿块和不对称致密。在使用了BREAST测试集进行训练后, 越南放射科医师在诊断灵敏度和毛刺或星芒状肿块检出率方面的得分提高了6%~30%^[29]。另一项研究^[30]也显示, 接受了BREAST培训的放射科医师比没有接受过培训的医师拥有更低的假阳性率。此外, Rawashdeh等^[20]的研究指出, 放射科医师的诊断能力与他们的阅片量呈正相关。BREAST可以为放射科医师提供大量优质的X线片进行训练, 这有助于提高医师在乳腺癌筛查中的灵敏度和准确度。

4 优势、不足和展望

BREAST作为一种可以提供即时在线反馈的乳腺X线摄影教育培训工具, 在许多方面都发挥着重要作用。① 个人: 在乳腺癌筛查的临床实践中, 阴性患者的数量远远高于阳性患者。对于年轻医师来说, 短时间内无法接触和学习到更多癌症案例, 想要快速提高诊断能力非常困难。BREAST拥有大量乳腺X线摄影和断层融合图像, 涵盖了多种类型的乳腺病灶, 可以为年轻医师提供典型患者集进行学习。此外, BREAST还可以为测试者提供实时在线反馈, 通过比较正确和错误之处, 测试者可以对自身进行评估, 了解自己的弱势并进行针对性学习。② 科室: 借助BREAST的分析和反馈功能, 科室领导者可以

及时获知每名医师的诊断能力和表现,从而制订培训计划来提高医师们的诊断水平。③ 国家: BREAST可以协助临床审核来进行乳腺癌筛查计划的质量保障和反馈,并且可以作为毕业后继续医学教育平台,为经验较少的年轻医师和专门从事乳腺癌筛查的医师提供交互式学习和培训。

尽管BREAST拥有诸多益处,但也存在一些有待改进的地方。例如,测试集中包含的癌症患者的比例很高,并不能完全反映真实的筛查环境,因此对于诊断特异度的提高效果不明显^[23, 29]。其次, BREAST测试集的难度水平无法完全标准化。在实际筛查环境中,不同影像学表现的乳腺癌发生比例各不相同。例如,表现为毛刺或星芒状肿块的患者发生概率远高于结构扭曲和不对称^[32]。此外,不同类型的乳腺癌以及不同密度的乳腺组织对于诊断的难度也均不相同。例如,大部分医师对于结构扭曲的误诊率较高,对致密型乳腺的漏诊率也相对较高^[15, 34]。因此,为了尽可能统一测试集的难度,需要更多资深的放射科医师参与患者选择,并且准确估计每例患者的难度指标。

总而言之, BREAST是一个可以为放射诊断科医师提供自我评估和提升能力的学习工具。在未来,开发者们将逐步扩大BREAST测试集并加入最新的成像技术,同时融入病理学、组织学、肿瘤学等内容,竭力将BREAST建立成为一个多专业的综合培训平台^[29]。

[参 考 文 献]

- [1] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics, 2020 [J]. *CA Cancer J Clin*, 2020, 70(1): 7–30.
- [2] NIELL B L, FREER P E, WEINFURTNER R J, et al. Screening for breast cancer [J]. *Radiol Clin N Am*, 2017, 55(6): 1145–1162.
- [3] RAWASHDEH M A, BOURNE R M, RYAN E A, et al. Quantitative measures confirm the inverse relationship between lesion spiculation and detection of breast masses [J]. *Acad Radiol*, 2013, 20(5): 576–580.
- [4] MORRELL S, TAYLOR R, RODER D, et al. Mammography screening and breast cancer mortality in Australia: an aggregate cohort study [J]. *J Med Screen*, 2012, 19(1): 26–34.
- [5] ALAKHRAS M, BOURNE R, RICKARD M, et al. Digital tomosynthesis: a new future for breast imaging? [J]. *Clin*

- Radiol*, 2013, 68(5): e225–e236.
- [6] WOODARD D B, GELFAND A E, BARLOW W E, et al. Performance assessment for radiologists interpreting screening mammography [J]. *Stat Med*, 2007, 26(7): 1532–1551.
- [7] ONEGA T, SMITH M, MIGLIORETTI D L, et al. Radiologist agreement for mammographic recall by case difficulty and finding type [J]. *J Am Coll Radiol*, 2016, 13(11S): e72–e79.
- [8] BREM R F, TABÁR L, DUFFY S W, et al. Assessing improvement in detection of breast cancer with three-dimensional automated breast US in women with dense breast tissue: the somoInsight Study [J]. *Radiology*, 2015, 274(3): 663–673.
- [9] GANDOMKAR Z, TAY K, RYDER W, et al. iCAP: an individualized model combining gaze parameters and image-based features to predict radiologists' decisions while reading mammograms [J]. *IEEE Trans Med Imaging*, 2017, 36(5): 1066–1075.
- [10] JACKSON R L, DOUBLE C R, MUNRO H J, et al. Breast cancer diagnostic efficacy in a developing south-east Asian country [J]. *Asian Pac J Cancer Prev*, 2019, 20(3): 727–731.
- [11] TRIEU P D Y, PUSLEDNIK L, COLLEY B, et al. Interpretative characteristics and case features associated with the performances of radiologists in reading mammograms: a study from a non-screening population in Asia [J]. *Asia Pac J Clin Oncol*, 2021, 17(1): 139–148.
- [12] CARNEY P A, BOGART T A, GELLER B M, et al. Association between time spent interpreting, level of confidence, and accuracy of screening mammography [J]. *AJR Am J Roentgenol*, 2012, 198(4): 970–978.
- [13] MIGLIORETTI D L, GARD C C, CARNEY P A, et al. When radiologists perform best: the learning curve in screening mammogram interpretation [J]. *Radiology*, 2009, 253(3): 632–640.
- [14] TABA S T, HOSSAIN L, WILLIS K, et al. Social networks and expertise development for Australian breast radiologists [J]. *BMC Health Serv Res*, 2017, 17(1): 131.
- [15] SULEIMAN W I, MCENTEE M F, LEWIS S J, et al. In the digital era, architectural distortion remains a challenging radiological task [J]. *Clin Radiol*, 2016, 71(1): e35–e40.
- [16] DESTOUNIS S, JOHNSTON L, HIGHNAM R, et al. Using volumetric breast density to quantify the potential masking risk of mammographic density [J]. *AJR Am J Roentgenol*, 2017, 208(1): 222–227.
- [17] KLOMPENHOUWER E G, DUIJM L E, VOOGD A C, et al. Variations in screening outcome among pairs of screening radiologists at non-blinded double reading of screening mammograms: a population-based study [J]. *Eur Radiol*, 2014, 24(5): 1097–1104.
- [18] PAN H B, YANG T L, HSU G C, et al. Can missed breast cancer be recognized by regular peer auditing on screening mammography? [J]. *J Chin Med Assoc*, 2012, 75(9): 464–467.

- [19] LI T, TABA S T, KHONG P L, et al. Reading high breast density mammograms: differences in diagnostic Performance between radiologists from Hong Kong SAR/Guangdong province in China and Australia [J] . *Asian Pac J Cancer Prev*, 2020, 21(9): 2623–2629.
- [20] RAWASHDEH M, ABDELRAHMAN M, ZAITOUN M, et al. Assessment of Jordanian radiologist performance in the detection of breast cancers [J] . *Open J Med Imaging*, 2018, 8(3): 41–53.
- [21] BRENNAN P C, GANDOMKAR Z, EKPO E U, et al. Radiologists can detect the ‘gist’ of breast cancer before any overt signs of cancer appear [J] . *Sci Rep*, 2018, 8(1): 8717.
- [22] RODER D M, WARD G H, FARSHID G, et al. Influence of service characteristics on high priority performance indicators and standards in the BreastScreen Australia program [J] . *Asian Pac J Cancer Prev*, 2014, 15(14): 5901–5908.
- [23] TRIEU P D Y, TAPIA K, FRAZER H, et al. Improvement of cancer detection on mammograms via BREAST test sets [J] . *Acad Radiol*, 2019, 26(12): e341–e347.
- [24] SULEIMAN W I, RAWASHDEH M A, LEWIS S J, et al. Impact of breast reader assessment strategy on mammographic radiologists’ test reading performance [J] . *J Med Imaging Radiat Oncol*, 2016, 60(3): 352–358.
- [25] SCOTT H J, GALE A G. Breast screening: PERFORMS identifies key mammographic training needs [J] . *Br J Radiol*, 2006, 79 Spec No 2: S127–S133.
- [26] REED W M, LEE W B, CAWSON J N, et al. Malignancy detection in digital mammograms: important reader characteristics and required case numbers [J] . *Acad Radiol*, 2010, 17(11): 1409–1413.
- [27] RAWASHDEH M A, BOURNE R M, RYAN E A, et al. Quantitative measures confirm the inverse relationship between lesion spiculation and detection of breast masses [J] . *Acad Radiol*, 2013, 20(5): 576–580.
- [28] GAGLIARDI A R, HONEIN–ABOUHAIDAR G, STUART–MCEWAN T, et al. How do the characteristics of breast cancer diagnostic assessment programmes influence service delivery: a mixed methods study [J] . *Eur J Cancer Care (Engl)*, 2018, 27(2): e12727.
- [29] TRIEU P D Y, LEWIS S J, LI T, et al. Improving radiologist’s ability in identifying particular abnormal lesions on mammograms through training test set with immediate feedback [J] . *Sci Rep*, 2021, 11(1): 9899.
- [30] TRIEU P D Y, PUSLEDNIK L, COLLEY B, et al. Interpretative characteristics and case features associated with the performances of radiologists in reading mammograms: a study from a non–screening population in Asia [J] . *Asia Pac J Clin Oncol*, 2021, 17(1): 139–148.
- [31] SOH B P, LEE W B, MELLO–THOMS C, et al. Certain performance values arising from mammographic test set readings correlate well with clinical audit [J] . *J Med Imaging Radiat Oncol*, 2015, 59(4): 403–410.
- [32] TRIEU P D Y, LEWIS S J, LI T, et al. Reader characteristics and mammogram features associated with breast imaging reporting scores [J] . *Br J Radiol*, 2020, 93(1114): 20200363.
- [33] REED W M, LEE W B, CAWSON J N, et al. Malignancy detection in digital mammograms: important reader characteristics and required case numbers [J] . *Acad Radiol*, 2010, 17(11): 1409–1413.
- [34] BANIK S, RANGAYAN R M, DESAUTELS J E L. Detection of architectural distortion in prior mammograms [J] . *IEEE Trans Med Imaging*, 2011, 30(2): 279–294.

(收稿日期: 2022–01–22 修回日期: 2022–03–15)