

# 基于深度学习的人工智能技术在乳腺癌筛查及影像诊断中的应用进展

潘德润 秦耿耿 陈卫国\*

【摘要】 人工智能技术可使计算机模拟人类的思考过程和智能活动,在医学影像领域具有强大的图像处理和特征提取能力,应用基于人工智能的乳腺癌影像筛查能够在减轻放射科医生工作负担的同时提高乳腺癌筛查和诊断的准确性及敏感性。简要总结了目前常用的乳腺X线图像公共数据集和近期乳腺癌影像学人工智能辅助诊断竞赛情况,并对基于深度学习的人工智能技术在乳腺癌影像学筛查和诊断中的最新进展予以综述。

【关键词】 乳腺X线摄影;乳腺癌筛查;深度学习;人工智能辅助诊断

中图分类号:R737.9;R445;TP181 文献标志码:A

**Progress in the application of artificial intelligence based on deep learning in breast cancer screening and diagnosis** PAN Derun, QIN Genggeng, CHEN Weiguo. Department of Radiology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China

【Abstract】 Artificial intelligence technology enables computer to simulate human thinking processes and intelligent activities. It has powerful image processing and feature extraction capabilities in medical imaging. The application of artificial intelligence-based breast cancer imaging screening can improve the accuracy and sensitivity of breast cancer screening and diagnosis while reducing the workload of radiologists. This paper briefly summarized the common mammographic databases and recent competition of artificial intelligence aided diagnosis of breast cancer imaging. The latest progress of artificial intelligence technology based on deep learning in breast cancer imaging screening and diagnosis was reviewed.

【Keywords】 Mammography; Breast cancer screening; Deep learning; Artificial intelligence aids diagnosis

*Int J Med Radiol, 2019, 42(1): 12-15*

乳腺癌是全球女性最常见的癌症,也是女性癌症死亡的主要原因<sup>[1]</sup>。乳腺癌如能在早期发现并及时治疗,其治疗效果也是恶性肿瘤中最佳的<sup>[2]</sup>。乳腺X线摄影是乳腺癌筛查和诊断的主要手段,可协助临床医生早期发现并及时治疗乳腺癌,从而显著降低其死亡率<sup>[3-4]</sup>。乳腺X线影像诊断通常是通过放射科医生浏览每幅影像,从中识别出肿块、钙化、结构扭曲及非对称致密影等常见异常征象,同时还需要结合病人的临床信息,这要耗费医生大量的精力。2006年,Hinton等<sup>[5]</sup>提出了“深度学习”的概念,它源于对人工神经网络的研究,含有多个隐藏层的多层

感知器是一种深度学习结构,能够通过模拟人脑的层次结构将数据从低层传递到高层,为最终的表示引入更多的语义信息<sup>[6]</sup>。深度学习方法已在文本、语音识别及自然图像的分类、检测等方面取得了重大突破<sup>[7-8]</sup>。因此,针对医学影像(非自然图像)领域开发出基于深度学习的计算机辅助诊断(computer aided diagnosis, CAD)系统,为放射科医生提供诊断的第二意见,以提高诊断敏感性和特异性<sup>[9-10]</sup>,这对于乳腺癌的早期筛查及诊断尤为重要。

## 1 乳腺X线图像数据集

基于深度学习的计算机乳腺癌辅助检测/诊断系统对于数据的需求是很大的,在设计、评估和调整CAD系统的过程中,研究人员往往需要大量的数字化乳腺X线图像<sup>[11]</sup>,它们对于比较不同研究的结果也很重要<sup>[12-13]</sup>,公开的数据集可以为研究人员开发、评估和对比评价提供一个共同基础。

作者单位:南方医科大学南方医院放射科,广州 510515

通信作者:陈卫国, E-mail: chenweiguo1964@21cn.com

\* 审校者

基金项目:广东省省级科技计划项目(2015B020233008, 2015B020233002);广州市科技计划项目(201604020058);广东省医学科学技术研究基金项目(A2017496)

DOI:10.19300/j.2019.Z6394zt

**1.1 乳腺 X 线摄影数字化数据集** (digital database for screening mammography, DDSM) DDSM 是目前最大的乳腺 X 线图像公共数据集, 也最为常用<sup>[14]</sup>。该数据集是由南佛罗里达大学等机构收集了 1988—1999 年 2 620 例病例, 共 10 480 幅图像。数字化图像是通过扫描胶片获得的, 格式为无损 JPEG 格式, 图像大小为 3 000×4 800 像素, 分辨率为 42  $\mu\text{m}$ , 包括内外斜位 (medial lateral oblique view, MLO) 和头尾位 (cranio-caudal view, CC) 视图, 具有正常、良性和恶性病变的图像, 异常病例均含有像素水平的标签。它还包含每个兴趣区 (ROI) 的类型, 如钙化或肿块, 大多数乳腺 X 线图像只包含一个 ROI。DDSM 包含病人的年龄、乳腺密度和乳腺影像报告和数据系统 (breast imaging reporting and data system, BI-RADS) 注释, 图像注释中还包括病灶在像素水平上的边界。用户可在 DDSM 数据集的网页 (<http://marathon.csee.usf.edu/Mammography/Database.html>) 上使用搜索引擎定义乳腺密度、扫描设备、病变病理及 BI-RADS 评估等关键词来搜索符合标准的病例。DDSM 的主要用途是促进 CAD 算法之间的比较<sup>[15]</sup>。另外, 一些研究者<sup>[16-17]</sup>经过验证指出 DDSM 的精确度不适用于验证分割算法。

**1.2 INbreast 数据集** INbreast 是公开可用的全视野数字乳腺 X 线摄影 (full-field digital mammography, FFDM) 数据集 (<http://medicalresearch.inescporto.pt/breastresearch/GetINbreastDatabase.html>), 由波尔图 CHSJ 乳腺中心提供。该数据集包括筛查、诊断和随访病例的图像。图像于 2008 年 4 月—2010 年 7 月期间获得, 共 115 例病例, 其中 90 例有每侧乳腺 2 个体位 (MLO 和 CC) 的图像, 其余 25 例为乳腺切除术的女性, 仅有单侧乳腺的 2 个体位的图像, 共 410 幅图像。图像大小根据病人乳腺大小分为 3 328×4 084 或 2 560×3 328 像素, 分辨率为 70  $\mu\text{m}$ , 保存为 DICOM 格式, 同一病人图像之间的对应关系与随机生成的病人身份标识一起保存。该数据集包含正常图像, 含有肿块和钙化、结构扭曲、非对称致密影及多种异常征象的乳腺 X 线图像, 此外还提供了病人的年龄、家族史、乳腺密度和 BI-RADS 分类的信息。INbreast 数据集的优点是具备精确的轮廓注释, 便于对评估肿块形态的算法进行训练及验证。

**1.3 乳腺 X 线图像分析协会数据集** (mammographic image analysis society, MIAS) MIAS 的数字化图像数据集是最早的公共数据集 ([https://www.repository.](https://www.repository.cam.ac.uk/handle/1810/250394)

[cam.ac.uk/handle/1810/250394](https://www.repository.cam.ac.uk/handle/1810/250394))<sup>[18]</sup>, 目前仍被广泛应用于研究中。该数据集包含 161 例病例, 322 幅数字化 MLO 图像, 原始图像由分辨率为 50  $\mu\text{m}$  的胶片扫描获得, 然后经过缩小、剪裁后得到 1 024×1 024 像素的图像, 格式为 PNG。该数据集包含正常图像以及良性、恶性病变的多种征象图像, 具有毛刺肿块的图像在数据集中占很高的比例。

在这 3 种数据集中, DDSM 具有丰富的病例类型, 可以很好地训练深度学习模型, 但它的数字化图像是通过扫描胶片获得的, 会产生一些伪影, 且精确度不足<sup>[19]</sup>, 会将某些目标像素错误地分配给背景而影响准确度。而 INbreast 为 FFDM 数据集, 具有精确的轮廓注释, 但病例数量相对较少。MIAS 数据集的缺点是数字化图像的对比度分辨率较低。

## 2 深度学习乳腺癌影像筛查竞赛

最近, 一些研究人员和赛智生物网络等组织发起了乳腺数字化梦想挑战 (DREAM Challenge), 挑战分为 2 个项目<sup>[20]</sup>: ①开发一种可以分析数字乳腺 X 线图像的预测算法; ②开发能够同时分析数字乳腺 X 线图像和临床信息的预测算法。他们收集了超过 640 000 幅乳腺 X 线图像作为竞赛数据集, 并与病人的临床数据和标准结果关联, 挑战数据集不包含 ROI 的注释。挑战者利用数字化图像和临床数据集来开发预测模型, 提交的模型将通过验证数据集进行评分。该竞赛的获胜团队在 DDSM 和 DREAM 挑战数据集 ([https://www.synapse.org/Digital\\_Mammography\\_DREAM\\_Challenge](https://www.synapse.org/Digital_Mammography_DREAM_Challenge)) 上分多个阶段对深度卷积神经网络进行训练, 他们用改进后的视觉几何组 (visual geometry group, VGG) 网络在以病灶为中心的补丁中对检测器网络进行预训练, 然后以完全卷积的方式将其插入更大的网络中, 并以端到端的方式在完整图像上对网络进行微调。该团队的预测模型在 2 个挑战项目中分别达到了 80.3% 和 80.4% 的准确度。

该竞赛创造了一个数据访问简单的开放社区, 同时激励参赛者共同协作并实时共享结果, 从中选拔出性能优异的深度学习技术及分类器, 这无疑在很大程度上加快了深度学习技术在乳腺癌筛查领域的发展速度。

## 3 深度学习技术近期进展

随着深度学习的发展, 目前已经有许多研究者采用深度学习方法来构建新型 CAD 系统<sup>[21-22]</sup>。同机器学习方法一样, 深度学习也分为监督学习与非监

督学习。如卷积神经网络(convolutional neural network, CNN)就是一种监督学习下的深度学习模型,而深度置信网络(deep belief nets, DBN)是一种非监督学习下的深度学习模型。基于 CNN 的模型往往需要对 ROI 进行详细的注释<sup>[23-24]</sup>,这耗费了大量的时间和费用,在监督学习尤其是医学成像方面造成了很大的困难。因此,一些研究者对弱监督学习方法进行了研究<sup>[25-26]</sup>,如多示例学习(multiple instance learning, MIL),它仅需要研究者提供整幅图像的标签,大大减少了训练成本。

**3.1 CNN** CNN 是深度学习模型的代表性结构,也是目前深度学习的研究热点。它是一种前馈型人工神经网络<sup>[6]</sup>,具有多层网络结构,通常包含输入层、卷积层、激活函数、池化层和全连接层。

CNN 具有强大的特征提取能力,可以提取较高等级的特征。Jiao 等<sup>[6]</sup>在 2016 年开发了基于 CNN 的 CAD 系统,对乳腺癌的肿块进行了分类。它主要包含一个 CNN 和一个决策机制。在训练过程中,用 CNN 提取了高级别和中等级别的特征,组合后对模型进行训练,并且将 CNN 自动提取的强度信息与深度特征结合起来,从而更好地模拟医生的诊断过程,取得了较好的效果。

另一方面,一些研究者研发了基于 CNN 的变型技术。2017 年,Al-Masni 等<sup>[27]</sup>提出了一种以区域深度学习技术为基础的 CAD 系统,该技术是一种基于 ROI 的 CNN,称之为 YOLO(You Only Look Once)。YOLO 是一种代表性的端到端训练算法,它在具有 ROI 信息的乳腺 X 线图像数据集中进行训练,并直接优化检测性能。此外,YOLO 学习目标的概括性表示具有很强的通用性,可以同时检测多种目标。YOLO 还可以同时学习 ROI 和背景,因此他们提出的 CAD 系统可以在一个 CNN 中完成特征提取并对乳腺肿块检测和分类,是一种快速、准确的目标检测器<sup>[28]</sup>。

还有一些研究者提出了减弱对 ROI 注释依赖性的 CNN。Li 等<sup>[20]</sup>采用全卷积设计开发了一种对完整乳腺 X 线图像进行乳腺癌诊断的端到端训练算法。该算法完全使用 CNN,因此可以输入任意大小的图像。此外,它只需要在训练的第一阶段对病变进行注释,训练模型识别局部补丁后,就可以将完整图像分类网络的权重初始化,之后该模型即可迁移为一个全图分类器,可在没有 ROI 注释的情况下进行端到端训练,这大大减少了对病变注释的依赖性。

与以往的方法相比,这种设计简单且性能更加优越。

**3.2 MIL** MIL 是一种弱监督学习方法,它是监督学习的变体,将图像表示为多示例包,基于包成分标签进行分类<sup>[29-30]</sup>,这样避免了对详细的医学图像注释的需求,大大降低了训练成本。“包”含有可变数量的示例(补丁),即使只有其中一个示例属于阳性,包也会被标记为阳性。只有当包中所有示例均为阴性时才会被标记为阴性。MIL 的目标是仅基于训练阶段的包标签来分类未知的包或示例。

2016 年,Quelleg 等<sup>[31]</sup>提出了一种新型的乳腺 X 线图像计算机辅助检测与诊断系统,该系统依赖 MIL 范例,仅使用整体图像水平的标签。他们首先将乳腺自适应地分割为多个区域,然后从每个区域中提取检测到的病变特征并合并,进而把乳腺 X 线图像分类为正常或异常。2017 年,Choukroun 等<sup>[10]</sup>描述了弱监督学习的计算机辅助检测和诊断系统,通过一个建立于 MIL 范例上的新型深度学习框架解决了乳腺 X 线图像异常结果的检测和分类问题。他们首先将图像分解为多个示例,通过选择最高阳性概率的示例完成检测任务,另外每个示例通过预训练网络产生一个特征向量,然后预测其类别概率,最后将所有示例的概率汇总得出整个“包”的类别概率。该方法的特点是可以利用 MIL 自动发现乳腺 X 线图像中的判别性示例。该系统的结果可与在完全注释的数据集中进行训练的监督方法相媲美。Zhu 等<sup>[32]</sup>也提出一种端到端训练的深度 MIL 神经网络,用于在没有 ROI 注释的情况下对乳腺 X 线图像进行分类。其目标是预测整个乳腺 X 线图像中是否包含恶性肿块。该研究同时使用了 CNN 来高效地获取所有示例的特征。

**3.3 DBN** DBN 是根据生物神经网络的研究及浅层神经网络发展而来的,属于双向深度网络<sup>[33]</sup>。DBN 由多层神经元构成,其关键组成元件是受限玻尔兹曼机。2016 年,Dhungel 等<sup>[34]</sup>使用一种深度学习和随机森林识别器级联的算法来检测乳腺 X 线图像中的可疑区域,该算法的第一阶段由多尺度 DBN<sup>[5]</sup>级联高斯混合模型<sup>[35]</sup>组成,网络选出候选区域后由深度 CNN 进一步处理,筛选出的区域由随机森林分类器进行分析。2017 年,Al-Antari 等<sup>[36]</sup>提出了一种基于 DBN 进行乳腺癌诊断的 CAD 系统,该系统包括肿块自动检测、ROI 提取、特征提取和 DBN 分类模块,其目标是识别正常、良性和恶性乳腺组织。他们首先通过无监督学习对受限玻尔兹曼机进行预训



练,然后根据反向传播算法来调整权重并微调该网络。DBN 的优点是可以利用所有提取的特征,并从中选择突出特征,研究结果表明该网络比传统的 CAD 系统具有更高的准确度。

#### 4 小结

深度学习技术在乳腺癌影像筛查领域已经有了诸多进展。传统的 CAD 系统没有深度网络,系统性能较差,仅能提取浅层特征。而近期提出的深度学习在乳腺癌筛查的应用中已经能够逐渐接近经验丰富的放射科医师的诊断水平<sup>[37]</sup>。然而,目前的深度学习技术对 ROI 注释的依赖性还没有得到很好的解决,因此未来几年可能会有更多关于减弱对 ROI 注释需求的深度学习技术的研究。另一方面,对 MIL 和 DBN 领域的研究相对较少,尽管它们避免了注释 ROI 的繁琐工作,但其系统性能尚待提高。还有研究者<sup>[32, 34]</sup>将多种深度学习技术联合运用,对乳腺 X 线图像中可疑区域进行检测和分类,这在一定程度上改善了模型的性能。此外,还需创建大型公共可用的数据集,使其图像具有较高的精确度、分辨率及多样性的特点,这将有助于训练出性能优越的 CAD 模型。

#### 参考文献:

- [1] Torre LA, Islami F, Siegel RL, et al. Global cancer in women: burden and trends[J]. *Cancer Epidemiol Biomarkers Prev*, 2017, 26: 444–457.
- [2] Migowski A. Early detection of breast cancer and the interpretation of results of survival studies[J]. *Cien Saude Colet*, 2015, 20: 1309.
- [3] Oeffinger KC, Fontham ETH, Etzioni R, et al. Breast cancer screening for women at average risk[J]. *JAMA*, 2015, 314: 1599–1614.
- [4] Weedon-Fekjaer H, Romundstad PR, Vatten LJ. Modern mammography screening and breast cancer mortality: population study[J]. *BMJ*, 2014, 348: g3701–g3708.
- [5] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313: 504–507.
- [6] Jiao Z, Gao X, Wang Y, et al. A deep feature based framework for breast masses classification[J]. *Neurocomputing*, 2016, 197: 221–231.
- [7] Zhang XL, Wang DL. Boosting contextual information for deep neural network based voice activity detection[J]. *IEEE/ACM Trans Audio Speech Lang Process*, 2016, 24: 252–264.
- [8] He T, Huang W, Qiao Y, et al. Text-attentional convolutional neural network for scene text detection[J]. *IEEE Trans Image Process*, 2016, 25: 2529–2541.
- [9] Giger ML. Medical imaging and computers in the diagnosis of breast cancer[C]. San Diego: Proc SPIE Int Soc Opt Eng, 2014.
- [10] Choukroun Y, Bakalo R, Benari R, et al. Mammogram classification and abnormality detection from nonlocal labels using deep multiple instance neural network[J]. *Eurographics Workshop Vis Comput Biomed*, 2017, 17: 11–19.
- [11] Nishikawa RM. Mammographic databases[J]. *Breast Dis*, 1998, 10: 137–150.
- [12] Garcia-Orellana CJ, Gallardo-Caballero R, Gonzalez-Velasco HM, et al. Study of a mammographic CAD performance dependence on the considered mammogram set[J]. *Conf Proc IEEE Eng Med Biol Soc*, 2008, 2008: 4776–4779.
- [13] Antoniou ZC, Giannakopoulou GP, Andreadis II, et al. A web-accessible mammographic image database dedicated to combined training and evaluation of radiologists and machines[C]. *Larnaca: IEEE*, 2009.
- [14] Heath M, Bowyer K, Kopans D, et al. Current status of the digital database for screening mammography[M]//Karssemeijer N, Thijssen M, Hendriks J, et al. *Digital mammography: Nijmegen*, 1998. Dordrecht: Springer, 1998: 457–460.
- [15] Heath M, Bowyer K, Kopans D, et al. The Digital database for screening mammography[M]//Netherlands: Springer, 2001: 457–460.
- [16] Sampat MP, Bovik AC, Whitman GJ, et al. A model-based framework for the detection of spiculated masses on mammography[J]. *Med Phys*, 2008, 35: 2110–2123.
- [17] Song E, Xu S, Xu X, et al. Hybrid segmentation of mass in mammograms using template matching and dynamic programming[J]. *Acad Radiol*, 2010, 17: 1414–1424.
- [18] Suckling J, Parker J, Dance DR. The mammographic image analysis society digital mammogram database [DB/OL]. <http://peipa.essex.ac.uk/info/mias.html>.
- [19] Singh S, Bovis K. An evaluation of contrast enhancement techniques for mammographic breast masses[J]. *IEEE Trans Inf Technol Biomed*, 2005, 9: 109–119.
- [20] Li S. End-to-end training for whole image breast cancer diagnosis using an all convolutional design [J/OL]. arXiv, 2017; arXiv:1711.05775 [2017–11–15]. <https://arxiv.org/ftp/arxiv/papers/1711/1711.05775.pdf>.
- [21] Arevalo J, Gonzalez FA, Ramos-Pollan R, et al. Convolutional neural networks for mammography mass lesion classification[C]. Milan: IEEE, 2015.
- [22] Chougrad H, Zouaki H, Alheyane O. Deep Convolutional Neural Networks for breast cancer screening[J]. *Comput Methods Programs Biomed*, 2018, 157: 19–30.
- [23] Akselrod-Ballin A, Karlinsky L, Alpert S, et al. A region based convolutional network for tumor detection and classification in breast mammography[C]. Cham: Springer International Publishing, 2016.
- [24] Becker AS, Marcon M, Ghafoor S, et al. Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer[J]. *Invest Radiol*, 2017, 52: 434–440.
- [25] Tang P, Wang X, Huang Z, et al. Deep patch learning for weakly supervised object classification and discovery[J]. *Pattern Recognit*, 2017, 71: 446–459.
- [26] Quellec G, Lamard M, Cozic M, et al. Multiple-instance learning for anomaly detection in digital mammography[J]. *IEEE Trans Med Imaging*, 2016, 35: 1604–1614.
- [27] Al-Masni MA, Al-Antari MA, Park JM, et al. Detection and classification of the breast abnormalities in digital mammograms via regional Convolutional Neural Network[C]. *Seogwipo: IEEE*, 2017.
- [28] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]. Las Vegas: IEEE, 2016.
- [29] Maron O, Lozano-Pérez T. A framework for multiple-instance learning[J]. *Adv Neural Inf Process Syst*, 1998, 200: 570–576.
- [30] Dieterich TG, Lathrop RH, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles[J]. *Artif Intell*, 1997, 89: 31–71.
- [31] Quellec G, Lamard M, Cozic M, et al. Multiple-instance learning for anomaly detection in digital mammography[J]. *IEEE Trans Med Imaging*, 2016, 35: 1604–1614.
- [32] Zhu W, Lou Q, Vang YS, et al. Deep multi-instance networks with sparse label assignment for whole mammogram classification[C]. Cham: Springer International Publishing, 2017.
- [33] Hinton GE, Osindero S, Teh Y. A fast learning algorithm for deep belief nets[J]. *Neural Comput*, 2006, 18: 1527–1554.
- [34] Dhungel N, Carneiro G, Bradley AP. Automated mass detection in mammograms using cascaded deep learning and random forests[C]. Adelaide: IEEE, 2015.
- [35] Biswas SK, Mukherjee DP. Recognizing architectural distortion in mammogram: a multiscale texture modeling approach with GMM[J]. *IEEE Trans Biomed Eng*, 2011, 58: 2023–2030.
- [36] Al-Antari MA, Al-Masni MA, Park SU, et al. An Automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network[J]. *J Med Biol Eng*, 2018, 38: 443–456.
- [37] Kooi T, Gubern-Merida A, Mordang J, et al. A Comparison between a deep convolutional neural network and radiologists for classifying regions of interest in mammography[C]. Cham: Springer International Publishing, 2016.

(收稿 2018–05–16)