

검진 유방촬영술에서의 인공지능 기반 진단보조프로그램의 적용

이시은 · 김은경

연세대학교 의과대학 용인세브란스병원 영상의학교실

서론

유방촬영술은 대규모 집단연구에서 사망률 감소 효과가 입증된 유일한 유방암 검진 방법이나, 유방암 발견에 대한 민감도와 특이도는 판독자에 따라 큰 차이를 보인다. 이는 유방암의 조기 발견을 늦추고, 때로는 불필요한 재검과 조직검사를 야기한다. 이를 개선하기 위한 노력으로 유방촬영술에서는 전통적 진단보조프로그램이 지난 10년 이상 사용되어 왔고, 최근에는 빅데이터에 기반한 기계학습의 발달에 힘입어 보다 정확한 유방암 진단을 돕기 위한 인공지능 알고리즘이 개발되어 적용되고 있다. 비록 현재까지는 후향적 분석에 기반하고 있으나, 다기관 연구를 통해 검증된 성능을 바탕으로 일부는 상용화되어 임상 환경에서 활용 범위를 넓혀 가고 있다 (1).

유방암 진단의 정확도를 높여줌과 동시에, 영상의학과 의사가 일상적으로 부담하고 있는 많은 양의 유방촬영 영상의 중요도를 분류함으로써 업무 부담을 경감하고, 나아가 인공지능이 예측한 위험도와 기존의 알려진 유방암 위험 인자를 종합한 모형을 개발, 유방암에 대한 위험도를 보다 세분화하고 이에 따른 환자 맞춤형의 검진 방식을 제안하는 데에도 기여할 수 있는 가능성을 보여주고 있다.

이번 종설에서는 검진 유방촬영술에 대해 인공지능 기반 진단프로그램 (이하 AI-CAD, artificial intelligence-based computer assisted diagnosis)을 적용한 주요 연구 결과를 정리해보고 이에 대한 임상 활용과 발전 과제에 대해 고찰해보고자 한다.

1. 판독 정확도의 향상

최근 몇 년 사이에 제안된 완성도 높은 AI 알고리즘은 유방암 진단에 있어 높은 정확도를 보여주고 있으며, 그 중 일부는 상용화되어 실제 판독에 활용되고 있다. 국내에서 가장 널리 사용되는 AI-CAD는 2020년 발표된 Lunit insight mammography이고, 한국, 미국, 영국에서 수집된 17만 장의 유방촬영 영상으로 학습한 결과, 14명 영상의학과 의사의 판독 정확도를 유의하게 향상시켰으며 이는 유방 영상을 전공하지 않은 영상의학과 의사들에서 더욱 뚜렷하게 나타났다 (2). AI-CAD는 과도한 위양성 결과가 가장 큰 문제였던 기존 전통적 CAD에 비해 이를 매우 낮게 유지하면서도 암에 대한 진단 정확도를 높이는 결과를 보였다 (Fig. 1). 자체 개발 알고리즘 (3) 및 다른 상용화 AI-CAD (4) 역시도 영상의학과 의사의 진단 정확도를 높임과 동시에, 단독 사용으로도 높은 정확도를 보였다. 다만 AI-CAD 프로그램은 학습된 데이터의 양과 질 및 기반 알고리즘의 차이에 따라, 상용화된 프로그램 사이에서도 그 성능에 차이를 보일 수 있다 (5).

현재까지 발표된 대부분의 성능 평가는 일반 검진 군에 비해서 상대적으로 암의 비중이 높은 집단에서 이루어져왔는데, 2021년 발표된 클라우드 소싱 형식의 DM DREAM (Digital Mammography - Dialogue on Reverse Engineering Assessment and Methods) challenge는 검진 유방촬영술에 대해 현재까지 가장 대규모로 검증한 결과를 보여준다 (6). 44개 국가의 126개 팀이 참여하여 경합하였고 10만장 이상의 검진 유방촬영술 영상으로 학습 및 20만 장의 영상으로 검증한 결과를 발표하였다. 이 중 어떤 단독 알고리즘도 영상의학과 의사 이상의 성능을 보이지는 못하였으나, 가장 높은 성능을 보인 8개 알고리즘의 조합과 영상의학과 의사의 판독 결과를 종합하였을 때에는 의사 단독의 결과에 비해 향상되는

통신저자: Eun-Kyung Kim, M.D., Ph.D.
Department of Radiology, Yongin Severance
Hospital Yonsei University College of Medicine 363,
Dongbaekjukjeon-daero, Giheung-gu, Yongin-si,
Gyeonggi-do, Korea
Tel. (031) 5189-8321 Fax. (02) 2227-8337
E-mail: ekkim@yuhs.ac

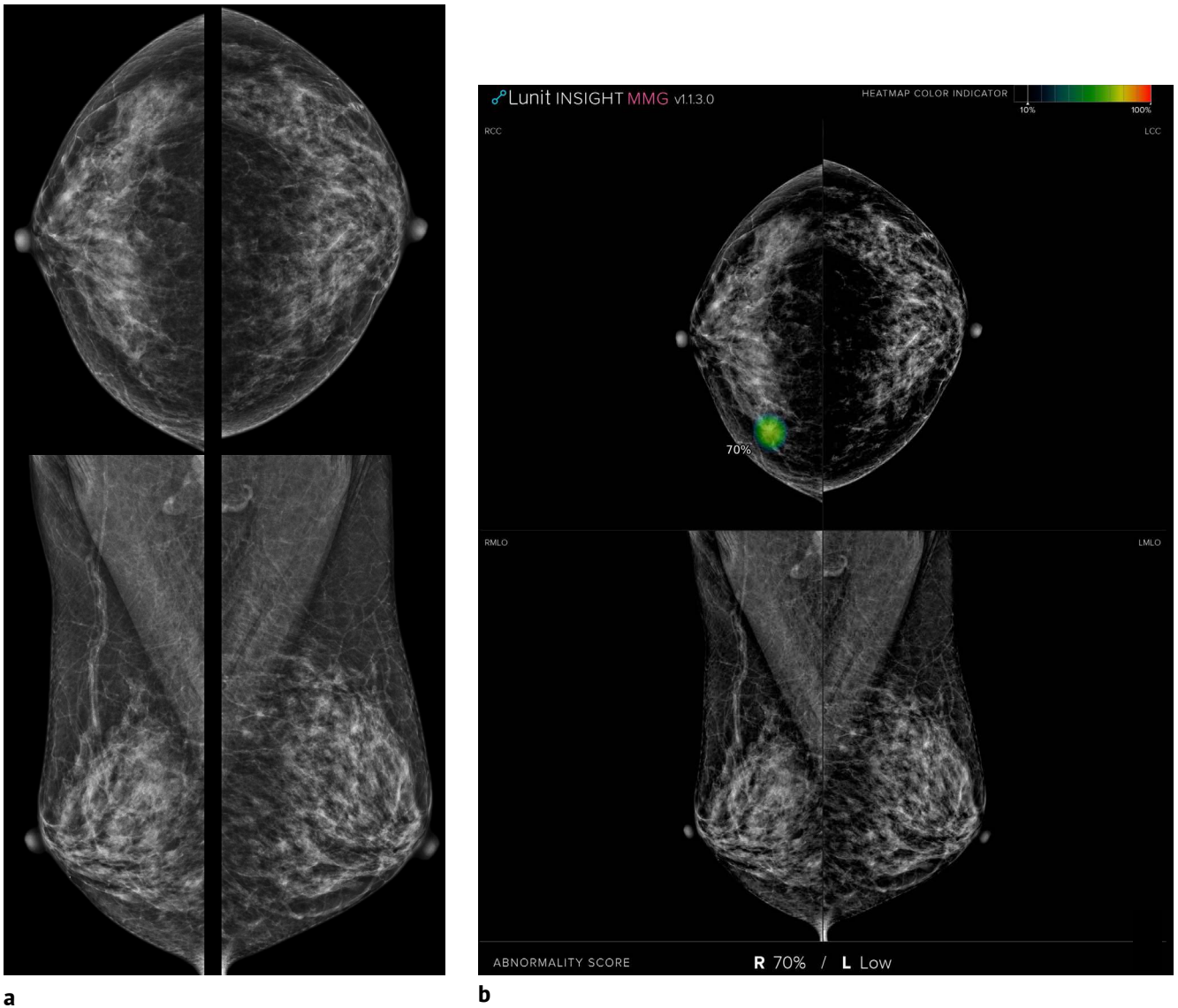


Fig. 1. A 59-year-old patient underwent screening mammography (a) which presented subtle distortion on right inner breast, detected by AI-CAD (b). It matched a 7 mm-sized suspicious nodule at 2 o'clock of right breast on US examination (c), which was invasive ductal carcinoma on US-guided biopsy and subsequent surgery.

결과를 보였다 (6). 기존 연구 결과가 AI-CAD 단독 사용으로도 영상의학과 의사에 비해 유의하게 높은 진단 정확도를 보여주었던 것에 비해, 대규모 검진 코호트에서 검증된 본 결과는 다소 기대에 미치지 못하였으나 이는 앞으로 발표될 전향적 연구를 통해 검증함이 바람직할 것으로 보인다.

2. 판독 효율성의 향상

일반적인 검진 유방촬영에서는 1% 미만이 유방암으로 진단되며, 대부분의 판독 업무는 정상 소견의 유방촬영술을 판독하는 데에 치중되어 있다. 검진의 활성화에 따라 유방촬영 건수는 지속적 증가추세에 있고, 판독 인력은 한정적이다. 유럽 지역에서는 한 유방촬영 영상에 대한 두 명의 영상의학과 의사에 의한 판독 시스템이 표준화되어 있으며, 이 경우 판독에 소요되는 시간은 더욱 길어질 수밖에 없다.

Lang 등은 대규모 검진 코호트에 상용 AI-CAD를 적용하였을 때, 설정한 역치점에 따라 암을 놓치지 않으면서도 정상 유방을 걸러낼 수 있음을 확인하였다 (7). 2012년부터 2020년까지 이루어진 AI-CAD를 활용한 환자 분류에 관한 7개 논문을 메타분석 하였을 때, 0-7%에 해당하는 유방암을 놓치는 범위에서 17-91%의 검사에 대한 판독을 정상 소견으로 대체 할 수 있었다 (8). 그러나 현재까지 발표된 AI-CAD의 위음성률은 20% 수준으로, 무시할 만한 결과는 아니다 (9). 다만 AI-CAD가 놓친 암은 환자의 예후에 영향이 적은, 검진으로 발견된 비교적 양호한 조직병리학적 특징을 가지는 병변 이었다는 점은 주목할 만하다 (9). AI-CAD에서 놓친 암은 영상의학과 의사에서도 판단이 어려운 병변인 경우가 많지만, 때로는 당황스럽게도 아주 뚜렷한 병변으로 나타나기도 한다 (Fig. 2). 이것은 AI-CAD가 위양성 결과를 줄이기 위해, 병변의 검출에만 집중하지 않고 검출된 병변의 악성도를 학습하고 평가한 결과를 나타내기 때문이며, 이에 대해서는 양성의 형태에 가까운 악성 병변에 대한 보다 정교한 학습이 필요한 부분으로 보인다. 이러한 점을 고려할 때, 현재 국내 실제 임상 상황에서 영상의학과 의사가 촬영된 영상을 확인하지 않고 AI-CAD 단독의 정상 결과를 보고하기는 어렵다. 그러나 두 명의 영상의학과 의사에 의한 이중 판독 시스템이 표준화된 국가에서는 한 명의 영상의학과 의사에 대한 보조 판독자로서 충분한 가치가 있을 것으로 여겨진다 (10).

AI-CAD로 정상 판독을 대체하지 않더라도, 이에 대한 결과를 참조하는 것 역시도 업무 흐름의 효율화를 가져올



Fig. 2. A 44-year-old patient diagnosed with a 20 mm-sized invasive ductal carcinoma in her left breast. Left breast cancer at 1 o'clock presenting as mass (arrow) was missed by AI-CAD.

수 있으며 이를 위해서는 PACS (Picture Archiving and Communication System)로 AI-CAD 분석 결과를 포함시켜 참조하기 편리한 환경을 구축하는 것이 필수적이다. 이를 통해 추가적으로 AI-CAD 결과를 확인하는데 소요되는 시간을 최소화하고, 유방 촬영 영상을 열어 확인하기 전 AI-CAD의 결과를 판독 리스트에 표기해줌으로써 판독해야 할 영상의 응급도/중요도를 분류할 수 있다 (Fig. 3). 일반적으로 판독에 소요되는 시간은 의심스러운 병변이 포함된 영상에서 훨씬 더 오래 소요되므로, 제한된 시간을 비정상 영상에 좀 더 효율적으로 사용할 수 있고 추가 검사를 요하는 환자의 판독을 가장 우선적으로 처리할 수 있게 된다.

3. 유방암 위험도 예측

초기 유방암 위험도 예측 모형은 가족력, 호르몬 인자 등 임상 인자에 기반하였으나, 유방촬영술 상의 치밀도가 중요한 위험 인자로 밝혀짐에 따라 유방암에 대한 예측력을 높였다 (11). 유방촬영술 상의 치밀도는 BI-RADS (Breast Imaging - Reporting and Data system) 기준에 따라 영상의학과 의사의 주관적 판단으로 4단계로 분류되고 있으나, 판독자간, 심지어는 한 판독자의 판독 내에서도 일치도가 낮은 것으로 알려져 있다 (12). 이를 개선하고자 전체 유방 면적 당 유선 조직 면적의 비율을 계산하는 프로그램도 활용되고 있으며, 최근에는 치밀도 판정을 위한 AI 알고리즘이 개발되어 영상의학과 의사의 판정과 높은 일치도를 보여준 바 있다 (13). 최신 버전의 Lunit insight mammography는 악성도 진단을 위한 기존 AI-CAD에 유방의 치밀도 정보를 추가하여 함께 제공하고 있

다 (14). AI-CAD는 유방 치밀도에 대한 보다 일관적인 평가를 제공할 뿐만 아니라, 영상의 픽셀 기반 정보에 기반하는 치밀도 이상의 유방암 위험도 평가에도 활용될 수 있다 (1).

최근 Yala 등이 발표한 유방암 위험도 예측 AI-CAD 프로그램은 임상적인 위험 인자에 기반한 기존의 위험도 예측 모형 (Tyrer-Cuzick model version 8)에 비해 더 높은 유방암 예측력을 보여주었다 (15). 또한, 예측된 위험도를 바탕으로 다음 촬영의 시기를 6-24개월로 조정하는 알고리즘을 구성하였을 때, 후향적 다기관 검증 결과 일률적으로 매년 시행하는 검진 유방촬영술에 비해 0.3-3개월의 유방암 진단 단축 효과를 보여줌과 동시에 전체적인 촬영 건수를 감소시켰다 (16). 전통적인 검진 방법의 검증 시 사용되는 사망률 감소 혹은 진행된 유방암이나 간격 유방암의 감소에 대한 분석은 이루어지지 않았으나, 일반적으로 유방 치밀도 외에 영상 조건을 전혀 포함하지 않았던

등록번호	이름	성...	SCORI...	검사명	검사날짜
97...	신...	F	97.33	Mammography (Both)	2022-02-03 10:19:12
95...	김...	F	87.94	Mammography (Both)	2022-02-03 14:52:19
98...	리...	F	66.28	Mammography (Both)	2022-02-04 15:45:50
95...	공...	F	59.05	Mammography (Both)	2022-02-04 09:09:57
64...	주...	F	55.91	Mammography (Both)	2022-02-03 11:21:38
98...	염...	F	49.66	Mammography (Both)	2022-02-03 13:52:26
95...	김...	F	38.3	Mammography (Both)	2022-02-04 08:46:07
32...	이...	F	36.08	Mammography (Both)	2022-02-04 08:57:32
99...	정...	F	18.66	Mammography (Both)	2022-02-03 10:48:27
98...	박...	F	15.99	Mammography (Both)	2022-02-04 09:51:34
99...	이...	F	9.9	Mammography (Rt)	2022-02-04 14:41:01
99...	송...	F	8.28	Mammography (Both)	2022-02-04 16:25:27
97...	정...	F	2.96	Mammography (Both)	2022-02-03 09:57:16
98...	강...	F	2.07	Mammography (Both)	2022-02-03 09:48:26
97...	김...	F	1.56	Mammography (Both)	2022-02-03 16:14:27
98...	차...	F	1.53	Mammography (Both)	2022-02-04 10:21:20
99...	김...	F	1.44	Mammography (Both)	2022-02-04 15:05:15
98...	이...	F	1.37	Mammography (Both)	2022-02-03 09:07:38
99...	백...	F	1.25	Mammography (Both)	2022-02-04 11:10:05
99...	오...	F	1.24	Mammography (Both)	2022-02-03 10:37:54
98...	심...	F	0.88	Mammography (Both)	2022-02-03 10:07:51
99...	전...	F	0.79	Mammography (Lt)	2022-02-04 14:10:19
99...	권...	F	0.64	Mammography (Both)	2022-02-03 15:58:50

Fig. 3. The AI-CAD is embedded in the PACS (Picture Archiving and Communication System). Before opening the examination, radiologists can refer the result on worklist as abnormality score (square) which can be sorted in ascending or descending power.

위험도 예측 모델에 AI-CAD가 이를 대표하는 정량화된 값으로 유용하게 활용될 수 있음을 보여주었다. 유방암의 긴 잠복기를 고려한다면, 유방암 발생 위험도의 예측은 현재 정상인 유방에서 앞으로 발생할 수 있는 유방암의 가능성 추정과 육안상 미묘한 병변의 이른 발견을 모두 포함할 수 있다. 복잡적이겠으나 후자의 경우 고위험 여성에서는 다음 유방 촬영의 시기를 당기는 것 보다는 추가적인 초음파나 MRI 검사를 권유하는 것이 더욱 유용할 수 있다. AI-CAD와 기존 임상적 위험인자를 병합한 예측 모형을 통해서 보다 정확히 유방암에 대한 위험도를 분류하고, 이에 맞춤형 검진 전략을 수립하는 것은 매우 중요한 주제이며, 여전히 많은 연구가 필요한 부분이다.

결론

최근 발전된 AI-CAD는 검진 유방촬영술의 판독 정확도를 높이고, 보다 효율적인 판독을 가능하게 한다. 또한 환자 중심의 보다 세분화된 위험도 평가에 기반한 유방암 검진을 향한 가능성을 열어주고 있다. 전향적 연구, 일반 검진 군을 대상으로 한 대규모의 외부 검증, AI-CAD에 대한 비용 산정 및 결과에 대한 책임 문제 등 AI-CAD의 활용 범위를 넓히기 위한 단계가 여전히 남아있으나, 감내할 만한 위양성 수준에서 유방암 발견율을 높일 수 있다는 점은 실제 판독 환경에서 보조 판독자로 활용하기에 충분한 이점으로 생각된다. AI-CAD의 발달은 현재 진행형이며, 영상 데이터를 통해 가치 있는 결과를 얻어냄으로써 보다 합리적이고 정확한 유방암 검진에 더욱 기여할 수 있기를 기대한다.

참고문헌

1. Yoon JH, Kim EK. Deep Learning-Based Artificial Intelligence for Mammography. *Korean J Radiol* 2021;22:1225-1239
2. Kim HE, Kim HH, Han BK, Kim KH, Han K, Nam H, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Health* 2020;2:e138-e148
3. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89-94
4. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. *J Natl Cancer Inst* 2019;111:916-922
5. Salim M, Wahlin E, Dembrower K, Azavedo E, Foukakis T, Liu Y, et al. External Evaluation of 3 Commercial Artificial Intelligence Algorithms for Independent Assessment of Screening Mammograms. *JAMA Oncol* 2020;6:1581-1588
6. Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. *JAMA Network Open* 2020;3:e200265-e200265
7. Lång K, Dustler M, Dahlblom V, Åkesson A, Andersson I, Zackrisson S. Identifying normal mammograms in a large screening population using artificial intelligence. *European Radiology* 2021;31:1687-1692
8. Hickman SE, Woitek R, Le EPV, Im YR, Luxhøj CM, Aviles-Rivero AI, et al. Machine Learning for Workflow Applications in Screening Mammography: Systematic Review and Meta-Analysis. *Radiology* 2022;302:88-104
9. Larsen M, Aglen CF, Lee CI, Hoff SR, Lund-Hanssen H, Lång K, et al. Artificial Intelligence Evaluation of 122 969 Mammography Examinations from a Population-based Screening Program. *Radiology* 2022;212381
10. Balta C, Rodriguez-Ruiz A, Mieskes C, Karssemeijer N, Heywang-Köbrunner S. Going from double to single reading for screening exams labeled as likely normal by AI: what is the impact?: SPIE, 2020
11. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;356:227-236
12. Sprague BL, Conant EF, Onega T, Garcia MP, Beaber EF, Herschorn SD, et al. Variation in Mammographic Breast Density Assessments Among Radiologists in Clinical Practice: A Multicenter Observational Study. *Ann Intern Med* 2016;165:457-464
13. Lehman CD, Yala A, Schuster T, Dontchos B, Bahl M, Swanson K, et al. Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. *Radiology* 2019;290:52-58
14. Lee SE, Son NH, Kim MH, Kim EK. Mammographic

- Density Assessment by Artificial Intelligence-Based Computer-Assisted Diagnosis: A Comparison with Automated Volumetric Assessment. *J Digit Imaging* 2022;35:173-179
15. Yala A, Mikhael PG, Strand F, Lin G, Satuluru S, Kim T, et al. Multi-Institutional Validation of a Mammography-Based Breast Cancer Risk Model. *J Clin Oncol* 2021;Jco2101337
16. Yala A, Mikhael PG, Lehman C, Lin G, Strand F, Wan YL, et al. Optimizing risk-based breast cancer screening policies with reinforcement learning. *Nat Med* 2022;28:136-143