

유방촬영술에서의 인공지능의 임상적용

장윤우

순천향대학교 서울병원 영상의학과

AI기술은 영상의학의 다양한 분야에서 매우 빠르게 발전하고 있으며 대표적으로 유방촬영술에서 유방암을 진단하는 인공지능 기반 진단보조 (AI-CAD) 프로그램이 상용화되어 있다. AI-CAD를 국가별로 다양한 유방촬영술 판독 시스템에 적용하거나 유방촬영술의 분류(triage)에 이용하여 유용성을 평가하는 후향적 분석이나 메타분석 연구들이 보고되었다. 이들 연구에서는 AI-CAD를 유방촬영술의 판독에 사용하면 유방암 발견의 판독 정확도와 진단 효율성을 향상시킨다고 보고하고 있으며 현재 몇 개의 대규모 전향적 연구들이 진행 중이다. 본 종설에서는 유방촬영술에 AI-CAD의 적용에 대해 정리해 보고 AI-CAD의 임상 적용에 대한 연구들과 국내에서 진행 중인 AI-STREAM연구를 포함하는 전향적 연구들에 대해 알아보려고 한다.

Index words: Artificial intelligence (AI); Mammography; Screening

유방촬영술의 인공지능 적용

인공지능(artificial intelligence, AI)이란 컴퓨터 과학을 통해 인공적으로 인간의 지능을 모방하여 구현하는 알고리즘 (algorithm)을 개발하는 것으로, 기계학습 (Machine learning)의 딥러닝 (Deep learning)등의 다양한 접근법을 포괄하는 용어이다. 영상의학 분야에서 AI 기술은 유방영상, 특히 유방촬영술 (mammography)을 이용한 연구가 많으며 매우 빠르게 증가하고 있다 (1-18).

유방촬영술에서 처음 사용되던 전통적인 컴퓨터보조진단 (conventional computed aided detection, CAD) 프로그램은 기계학습 (Machine learning) 중 지도학습 (supervised learning)을 통하여 개발되었는데 특정

한 컴퓨터 프로그래밍 없이 BI-RADS (Breast Imaging Reporting and Data System)에 입각하여 유방영상 전문가가 정의해 준 유방암의 특징들을 유방촬영술에서 암이 의심되는 위치에 표기하여 보여주는 방식이다. 유방촬영술에 대한 인공지능 기반 진단 보조 (AI-CAD) 프로그램은 기계학습 (Machine learning) 중 딥러닝 (deep learning) 알고리즘 방법을 사용하는데 인간의 신경망을 모방하여 연속적으로 쌓여 있는 합성곱 신경망 (convolutional neural network, CNN)을 사용하여 다양한 공간적 정보로부터 특정한 작업을 수행하는데 필요한 주요 정보들을 스스로 배울 수 있다고 알려져 있다. 이러한 딥러닝 (deep learning) 알고리즘을 사용하여 대규모 유방 영상 데이터에서 정상 (negative) 인지 이상소견이 있는지 (positive)로 분류 (classification)하는데 필요한 기능을 학습하여 raw data에서 점차 높은 수준의 정보를 추출하게 되는데 많은 데이터에 노출될수록 인간에게 알려지지 않았거나 인지하기 어려운 유용한 특징을 발견할 수 있다 (1, 2).

통신저자: Yun-Woo Chang, M.D., Ph.D.
Department of Radiology
Soonchunhyang University Hospital Seoul
59 Daesakwan-ro, Yongsan-ku, Seoul 04401, Korea
Tel. +82-2-709-9396 Fax. +82-2-709-3928
E-mail: ywchang@schmc.ac.kr

유방촬영술의 한계와 CAD의 사용

유방촬영술은 유방 검진 (screening) 시 가장 먼저 시행하는 검사법으로 유방암을 조기에 발견하여 유방암에 의한 사망률을 낮출 수 있는 것으로 검증된 유일한 검사법이다 (19). 우리나라에서는 1999년부터 유방촬영술을 이용한 국가암 검진 사업을 시행하여 만 40세이상 여성에서 2년마다 유방촬영 검사를 시행하고 있다(3). 그러나 유방촬영술은 다음과 같은 몇 가지 제한점을 가지는데 먼저 치밀 유방에서는 유방 실질에 병변이 가려지고 병변을 발견하는 민감도가 떨어질 수 있으며 양성과 악성을 구분하는데 제한이 있다. 또한, 이상소견을 인식하지 못하여 위음성으로 해석하는 경우 유방암 진단이 지연될 수 있으며, 추가적인 영상 검사나 조직 검사가 필요하다고 간주하는 경우 불필요한 위양성 회상 (recall)으로 이어질 수 있다 (20-22). 아시아여성에서는 치밀 유방의 비율의 높음에 우리나라 40대이상 국가암 검진 수검자의 약 70%가 치밀 유방으로 보고되었다 (3). 이런 이유로 유방촬영술은 높은 수준의 민감도를 확보하기 어려우며 판독에 있어서도 높은 전문성을 필요로 하는 검사이다. 이러한 유방촬영술의 문제점을 극복하기 위해 이중 판독 (double reading)을 수행하거나, 선별검사의 빈도를 높이거나, 추가적인 보충영상 검사 (예, 초음파, DBT (digital breast tomosynthesis), MRI 등)를 시행하여 상당한 개선을 보였다(20, 23). 검진을 강화하고 추가적인 보충 영상 기법을 사용하면 유방암 발견율이 향상될 수 있지만 영상의 해석 부담이 증가하고 불가피하게 의료비가 증가하게 되므로 충분한 자원 확보가 문제가 될 수 있다. 그러나 이러한 의료자원은 여전히 제한적이다 (20). DBT가 유방촬영술의 대안으로 사용될 수 있지만 여전히 치밀 유방 (dense breast)에서는 민감도 (sensitivity)가 떨어지고 위양성 회상 (recall)이 증가되며 판독자에 따른 판독 성능에 차이도 무시하기 어려운 문제점이다 (24).

유방암 발견의 정확도를 향상시키기 위한 목적으로 유방촬영술에서 유방암 의심 병변을 자동으로 검출하는 CAD 혹은 CADx (컴퓨터보조진단)이라는 알고리즘이 개발되어 1998년 처음 FDA허가를 받고 미국 병원들에 도입되었다. 이런 conventional CAD는 CAD없이 판독 후 CAD가 표시한 부분을 다시 한번 확인하는 방식으로 사용되었는데 초기의 연구들에서는 CAD를 앞서 언급된 second reader 역할로 사용하였을 때 ‘인정할 만하다’는 결과들이 보고되었다. 이를 토대로 미국에서는 보험 수가를 인정받고 90%이상 (2016년 기준)의 미국 병원들에

CAD가 보급되었다(2, 7, 8, 25). 그러나 높은 빈도의 위양성 판독으로 인해 한 개의 유방암을 찾기 위해 훨씬 더 많은 수의 위양성을 감수해야 하는 상황이 판독자로 하여금 피로감을 유발하여 실제적으로 참고를 하지 않게 되었고 대규모의 전향적 연구들에서도 CAD의 효용성을 확실하게 입증되지 못하였다 (4, 5, 7, 13, 26). 최근에는 유방촬영술에서 딥러닝 (deep learning)을 통한 AI-CAD에 대한 연구들이 매우 빠르게 증가하고 있다.

유방촬영술 판독 방식 (workflow)의 다양성과 AI-CAD의 적용

유방촬영술을 해석하는 방법은 국가별로 다양한 전략을 가지고 있으므로 나라마다 다양한 작업 흐름(workflow)에 AI-CAD를 적용하는 연구들이 보고되고 있다 (20). 우리나라나 미국 같이 영상의의 단독 판독 (single reading)을 하는 경우에는 AI-CAD의 도움을 추가하는 AI reader aid에 대한 연구들이 보고되었다 (9, 10, 12, 15). 유럽 국가들의 경우 두명의 의사가 이중 판독 (double reading)을 시행한 후 의견이 불일치하는 경우 합의 판독 (consensus reading)이나 중재판독 (arbitration reading)을 시행하는 것을 표준판독법 (standard method)으로 한다. 이러한 판독 전략에서는 AI-CAD가 하나의 reader 역할을 하거나 추가적 합의 혹은 중재 판독에 AI-CAD의 도움을 받는 경우에 대한 연구들이 보고되었다(11, 15, 16, 26, 27). 그 외에 AI-CAD를 사용해 효과적으로 유방촬영술을 분류 (triage)하여 업무량을 줄여보려는 연구들도 보고되고 있다 (28-30). 유방암 검진 시 발견되는 유방암의 비율이 1%에 미치지 못한다는 점을 감안하여 대부분의 유방촬영술 선별검사 결과가 음성이기 때문에 AI-CAD를 사용한 정확한 분류를 통해 환자의 위험을 증가시키지 않으면서 더 적은 시간과 리소스를 사용하여 영상의의 업무량을 줄일 수 있으며 영상의는 더 복잡하고 전문성이 요구되는 정밀 검사에 더 많은 시간을 투자할 수 있을 것이다.

딥러닝 AI 알고리즘을 실제 임상 진료에 사용하기 위해서는 다음의 단계들을 거쳐야 하는데 먼저 적절한 임상자료를 이용한 임상 검증 (clinical validation)이 이루어져야 한다. 내부 성능평가 (internal validation)를 통해 알고리즘 개발에 사용된 데이터 중 일부를 이용한 검증 이후 학습에 사용되지 않은 다양한 제조사 제품, 촬영 조건, 영상 후처리 방식 등의 다양한 디지털 영상자료를 이용하는 외부 검증 (external validation)을 통한 성능평가를 시행

하며 이와 함께 독자연구 (reader study)를 통하여 수백장의 유방 촬영술에서 AI를 사용하지 않은 결과와 사용한 결과를 비교하여 영상의(radiologist)의 판독 정확도 향상을 입증하는 평가가 필요하다. 이후 실제 후향적 코호트 연구를 통해 선택편향의 효과 (bias)를 최소화한 상태에서 실제 유방암 검진환경에서 진단능의 평가를 시행하여 임상적 가치에 대한 근거를 마련한다. 마지막 단계로 전향적 연구를 통한 유방암 검출율 (cancer detection rate) 및 회상율 (recall rate)의 변화를 실제 유방암 검진 코호트에서 입증하여야 한다(3). 현재까지의 유방촬영술에 대한 AI-CAD의 사용은 독자연구 (reader study) 나 후향적 연구 (retrospective study)들이 보고되고 있으며 AI-CAD를 사용하였을 때 유방촬영술의 판독성능이 향상을 보였고 AI-CAD의 단독 판독도 영상의의 단독 판독과 비교하여 열등하지 않거나 오히려 영상의의 판독값을 능가한다는 결과들이 보고되었다 (6, 9, 17, 24, 31). 하지만 이들 후향적인 분석의 연구 결과들을 뒷받침하기 위해서는 독립적인 전향적인 검증이 필요하다. 현재 Clinical trial에 보고된 몇 개의 전향적 연구들의 예비 결과가 보고되었다 (32, 33).

전향적 연구인 ScreenTrustCAD는 실제 의료환경에서 스웨덴 여성 55,581명을 대상으로 유럽 가이드라인에 따른 영상의 2명의 표준 이중 판독 결과를 영상의 1명과 AI-CAD (Lunit INSIGHT MMG, Lunit)를 적용한 결과 및 AI-CAD단독 결과와 비교하였다. 영상의 1명과 AI-CAD가 판독한 경우 표준 이중 판독 결과보다 암 발견율이 4% 더 높았고 AI-CAD 단독 판독의 경우도 표준 이중 판독에 비해 열등하지 않았다. 이와 함께 재검사를 위해 환자를 다시 소환하는 회상율을 비교하였을 때도 영상의 1명과 AI-CAD가 판독한 경우의 합의 판독의 회상율이 표준 이중 판독 후 합의 판독의 회상율보다 4% 더 낮았다 (33).

또다른 전향적 연구인 MASAI (Mammography Screening with Artificial Intelligence) 연구는 스웨덴에서 100만명을 모집하는 것으로 계획되어 진행되고 있는 단일센터 무작위 대조시험 (single center randomized controlled trial)으로 연구에 등록된 환자들은 무작위 배정을 통해 유럽의 표준 이중 판독을 하는 그룹과 AI-CAD를 적용하는 대조그룹으로 나누었다. 연구에 사용된 AI-CAD (Transpara, ScreenPoint Medical)의 위험 점수는 1-10 범위의 연속 척도로 제공되며 보정된 10단계 개별 척도로 표시되었다. 위험척도는 저위험 (위험점수 1-7), 중위험 (위험점수 8-9), 고위험 (위험점수 10)으로 간주하며 위험점수가 10인 그룹에서 암유병률이 급격하게 증가한

다. 이는 동일한 AI-CAD를 이용한 후향적 연구 결과와도 일치한다 (29, 32). MASAI연구에서는 AI-CAD를 적용하는 대조그룹에서는 score 1-9까지는 단독 판독을 시행하고 score 10인 경우만 이중 판독을 시행하여 (risk score는 제공), 회상 (recall) 여부를 판정하였을 때 암발견율은 표준 이중 판독에 비해 부족하지 않았고 회상율이나 위양성에는 영향을 미치지 않으면서 검진 판독량을 44% 감소시켰다고 하여 AI를 사용한 분류 (triage)를 고려할 수 있다는 예비 결과를 보고하였다(32).

AI-CAD의 전향적 구현에 대한 또다른 연구에서는 이중 판독 후 회상 (recall) 하지 않은 것으로 판정된 유방촬영술에 AI-CAD를 적용하여 AI 시스템이 이상이 있다고 표시한 사례에 대해 추가적인 영상의의 중재적 평가를 시행하는 방법을 제시하였다. 이 평가는 AI-CAD를 추가 판독기로 사용하면 불필요한 회상을 최소화하면서 유방암 조기 발견을 개선할 수 있음을 시사하였으며 전체 판독 작업량이 4-6%정도의 추가 읽기가 필요하지만 양성예측치 (PPV)가 증가를 보여 스크리닝의 효율성을 높일 수 있음을 보여주었다 (35).

이들 연구는 이중 판독 전략을 표준 판독법으로 하는 유럽 국가들에서 진행된 연구결과로 우리나라나 비유럽 국가들 같이 유방촬영술 단독 판독 (single reading)을 표준 판독법으로 하는 환경에서의 AI-CAD의 적용이 미치는 영향에 대한 전향적 연구가 필요하다. AI-STREAM (Artificial Intelligence for breast cancer screening in Mammography) 연구는 다기관 전향적 코호트 연구로 국가암 검진 유방촬영을 시행하고 있는 국내 6개 대학병원 (강동 경희대병원, 순천향대 서울병원, 노원 을지대병원, 건국대 병원, 인제대 부산백병원, 분당 차병원)에서 2021년 2월부터 2022년 12월까지 25,008명을 모집하였다. AI-stream연구의 목적은 AI-CAD (Lunit INSIGHT MMG, Lunit)를 사용하였을 때 실제 유방암 검진에서 유방암 발견율과 회상율에 미치는 영향을 알아보고 더 나아가서 발견된 유방암들의 특성 등을 전향적으로 검증하고자 하였다 (34). 유방촬영술 단독 판독에서 영상의가 AI-CAD를 사용한 경우와 AI-CAD를 사용하지 않은 경우를 비교하여 AI-CAD의 도움을 받은 경우 검진암 발견율을 높인다는 예비 분석 결과가 가까운 시일내에 발표될 예정이다. 2년의 추적 검사로 발견되는 간격암 (interval cancer)에 대한 최종 분석 결과는 2026년 이후 확인할 수 있을 것으로 보인다.

이들 전향적 연구를 통해 AI-CAD를 적용하였을때 국가별로 다양한 판독 전략을 가지는 유방촬영술의 판독에

서 암발견율이나 회상율에 어떤 긍정적인 영향을 미칠 수 있는지에 대한 근거가 마련될 것으로 보이며 실제 유방암 검진 환경에서 AI-CAD 사용의 임상적 가치를 입증할 수 있을 것으로 기대된다.

참 고 문 헌

1. Bahl M. Artificial Intelligence: A Primer for Breast Imaging Radiologists. *J Breast Imaging*. 2020;2(4):304-314.
2. Fujita H. AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. *Radiol Phys Technol*. 2020;13(1):6-19.
3. Kim KH, Lee SH. [Applications of Artificial Intelligence in Mammography from a Development and Validation Perspective]. *Taehan Yongsang Uihakhoe Chi*. 2021;82(1):12-28.
4. Cole EB, Zhang Z, Marques HS, Edward Hendrick R, Yaffe MJ, Pisano ED. Impact of computer-aided detection systems on radiologist accuracy with digital mammography. *AJR Am J Roentgenol*. 2014;203(4):909-916.
5. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med*. 2007;356(14):1399-1409.
6. Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A, et al. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *Bmj*. 2021;374:n1872.
7. Freer TW, Ullissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology*. 2001;220(3):781-786.
8. Gilbert FJ, Astley SM, Gillan MG, Agbaje OF, Wallis MG, James J, et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med*. 2008;359(16):1675-1684.
9. 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(3):e138-e148.
10. Kim YS, Jang MJ, Lee SH, Kim SY, Ha SM, Kwon BR, et al. Use of Artificial Intelligence for Reducing Unnecessary Recalls at Screening Mammography: A Simulation Study. *Korean J Radiol*. 2022;23(12):1241-1250.
11. Lång K, Dustler M, Dahlblom V, Åkesson A, Andersson I, Zackrisson S. Identifying normal mammograms in a large screening population using artificial intelligence. *Eur Radiol*. 2021;31(3):1687-1692.
12. Lee SE, Han K, Yoon JH, Youk JH, Kim EK. Depiction of breast cancers on digital mammograms by artificial intelligence-based computer-assisted diagnosis according to cancer characteristics. *Eur Radiol*. 2022;32(11):7400-7408.
13. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. *JAMA Intern Med*. 2015;175(11):1828-1837.
14. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, et al. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. *Radiology*. 2019;290(2):305-314.
15. Rodríguez-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(9):916-922.
16. Romero-Martín S, Elías-Cabot E, Raya-Povedano JL, Gubern-Mérida A, Rodríguez-Ruiz A, Álvarez-Benito M. Stand-Alone Use of Artificial Intelligence for Digital Mammography and Digital Breast Tomosynthesis Screening: A Retrospective Evaluation. *Radiology*. 2022;302(3):535-542.
17. 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 Netw Open*. 2020;3(3):e200265.
18. Watanabe AT, Lim V, Vu HX, Chim R, Weise E, Liu J, et al. Improved Cancer Detection Using Artificial Intelligence: a Retrospective Evaluation of Missed Cancers on Mammography. *J Digit Imaging*. 2019;32(4):625-637.
19. Myers ER, Moorman P, Gierisch JM, Havrilesky LJ, Grimm LJ, Ghatge S, et al. Benefits and Harms of

- Breast Cancer Screening: A Systematic Review. *Jama*. 2015;314(15):1615-1634.
20. Yoon JH, Kim EK. Deep Learning-Based Artificial Intelligence for Mammography. *Korean J Radiol*. 2021;22(8):1225-1239.
21. Hovda T, Tsuruda K, Hoff SR, Sahlberg KK, Hofvind S. Radiological review of prior screening mammograms of screen-detected breast cancer. *Eur Radiol*. 2021;31(4):2568-2579.
22. Lamb LR, Mohallem Fonseca M, Verma R, Seely JM. Missed Breast Cancer: Effects of Subconscious Bias and Lesion Characteristics. *Radiographics*. 2020;40(4):941-960.
23. Taylor-Phillips S, Stinton C. Double reading in breast cancer screening: considerations for policy-making. *Br J Radiol*. 2020;93(1106):20190610.
24. Yoon JH, Strand F, Baltzer PAT, Conant EF, Gilbert FJ, Lehman CD, et al. Standalone AI for Breast Cancer Detection at Screening Digital Mammography and Digital Breast Tomosynthesis: A Systematic Review and Meta-Analysis. *Radiology*. 2023;307(5):e222639.
25. Gao Y, Geras KJ, Lewin AA, Moy L. New Frontiers: An Update on Computer-Aided Diagnosis for Breast Imaging in the Age of Artificial Intelligence. *AJR Am J Roentgenol*. 2019;212(2):300-307.
26. 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(7788):89-94.
27. Salim M, Wåhlin 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(10):1581-1588.
28. Dembrower K, Wåhlin E, Liu Y, Salim M, Smith K, Lindholm P, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. *Lancet Digit Health*. 2020;2(9):e468-e474.
29. 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;303(3):502-511.
30. Leibig C, Brehmer M, Bunk S, Byng D, Pinker K, Umutlu L. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. *Lancet Digit Health*. 2022;4(7):e507-e519.
31. Hickman SE, Woitek R, Le EPV, Im YR, Mouritsen Luxhøj C, Aviles-Rivero AI, et al. Machine Learning for Workflow Applications in Screening Mammography: Systematic Review and Meta-Analysis. *Radiology*. 2022;302(1):88-104.
32. Lång K, Josefsson V, Larsson AM, Larsson S, Högborg C, Sartor H, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *Lancet Oncol*. 2023;24(8):936-944.
33. Dembrower K, Crippa A, Colón E, Eklund M, Strand F. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit Health*. 2023;5(10):e703-e711.
34. Chang YW, An JK, Choi N, Ko KH, Kim KH, Han K, et al. Artificial Intelligence for Breast Cancer Screening in Mammography (AI-STREAM): A Prospective Multicenter Study Design in Korea Using AI-Based CAdE/x. *J Breast Cancer*. 2022;25(1):57-68.
35. Ng AY, Oberije CJG, Ambrózy É, Szabó E, Serfőző O, Karpati E, et al. Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer. *Nat Med*. 2023;29(12):3044-3049.