

인공지능 기반 컴퓨터 보조진단을 이용한 선별 유방촬영술에서의 간격암에 대한 후향 분석

김연수 · 장정민

서울대학교 의과대학 서울대병원 영상의학과

목적: 선별 유방촬영술에서 발견되지 않은 간격암에 대하여 후향적으로 인공지능 기반의 컴퓨터 보조진단 진단능을 분석하고자 한다.

대상 및 방법: 2016년 1월부터 2019년 12월까지 새롭게 유방암을 진단받고 수술 전 유방촬영술을 시행한 환자 중 1년 이내 이전 선별 유방촬영술이 음성인 33명 환자 (40-76세, 평균 54.2세)의 33개의 유방을 대상으로 하였다. 두 명의 유방 영상 전문의가 합의하에 이전 유방촬영술을 재평가하여 33개의 간격암을 놓친암과 진성 간격암으로 분류하였고, 이를 인공지능 기반 컴퓨터 보조진단 (Artificial Intelligence based Computer Aided Diagnosis; AI-CAD)의 결과와 비교하였다.

결과: 33개의 간격암은 재평가시 소환 (recall)이 필요했던 (actionable) 12개 (36.4%)의 놓친암과 다시 보아도 병변이 보이지 않는 21개 (63.6%)의 진성 간격암으로 분류되었다. 그 중 AI-CAD는 영상에 36개의 표지를 하였고 그 중 13개 (36.1%)가 진양성 암에 표시되었다. AI-CAD는 전체암 중 9예 (27.3%, [9/33])를, 놓친암 중에서는 7예 (58.3%, [7/12])의 암을 추가적으로 찾았다.

결론: AI-CAD는 유방촬영술상 소환이 필요했던 놓친 암의 상당부분을 정확하게 진단하였다. 선별 유방촬영술에서 AI-CAD의 이용은 영상의학과 의사의 해석 오류로 인해 발생하는 놓친암을 줄일 수 있을 것으로 기대된다.

Index words: Breast; Mammography; Missed cancer; Artificial intelligence

서 론

유방촬영술은 무작위 대조군 연구를 통해 유방암으

로 인한 사망률을 감소시키는 것으로 입증된 유일한 유방암 선별검사법이다 (1). 유방촬영술에서 약 10-30%의 유방암을 놓칠 수 있는데 그 원인으로는 치밀한 섬유유선조직으로 인한 병변 가림, 영상의학과 의사의 인식 (detection) 혹은 해석 (interpretation) 오류, 부적절한 자세 잡기 등이 있다 (2). 유방촬영술은 판독자에 따라 영상에 대한 해석이 광범위하게 달라질 수 있어 그 진단율이 다르다 (3). 유방촬영술 판독에 있어 민감도가 증가되면 수반되는 위양성의 증가는 환자의 불안, 불필요한 검사 및

본 연구는 주식회사 동국생명과학(주)의 지원 (서울대병원 위탁과제 06-2020-2840) 을 받아 수행되었습니다.

통신저자: 장정민

(03080) 서울시 종로구 대학로 101, 서울대학교 의과대학
서울대병원 영상의학과

Tel. (02) 2072-3617, Fax. (02) 743-6385

E-mail: imchangjm@gmail.com

침습적인 검사를 증가시킨다.

1990년대에 처음 도입된 유방촬영술을 위한 컴퓨터 보조 진단 (computer aided diagnosis, CAD)은 민감도를 높여 진단능 향상에 도움을 주나, CAD가 보여주는 많은 위양성 표시에 대한 평가와 추가 검사에 대한 판단에 있어 진단 효율성에는 여전히 이견이 있는 상태이다 (4, 5). 최근 각광을 받고 있는 딥러닝 방식은 20여년 전부터 도입되어 온 컴퓨터 보조 진단의 특수한 형태 중 하나이다. 이 상적인 선별검사는 상대적으로 낮은 사회적 비용으로 높은 민감도와 암 발견율, 그리고 낮은 위양성을 보여야 한다 (3). 이에 선별 유방촬영술에 대한 인공지능 기반 컴퓨터 보조진단 (Artificial Intelligence based Computer Aided Diagnosis; AI-CAD)은 빠른 인식에 도움을 주고 해석에 있어서도 하나의 선택지로 작용할 수 있다는 점에서 큰 잠재력을 갖고 있으나 (6), 판독자의 감독 하에 선택적 적용을 필요로 한다.

이에 본 연구는 선별 유방촬영술에서 당시 판독상 음성이었던 놓친암을 포함한 간격암에 대해 AI-CAD를 적용해보고, 그 중에서 특히 놓친 암에 대한 원인을 분석하여 AI-CAD가 도움이 되는 적응증을 알아보고자 한다.

대상 및 방법

후향적 데이터를 이용한 본 연구는 임상 윤리 심의 위원회에서 승인을 받았다. 2016년 1월부터 2019년 12월까지 본원 및 강남건진센터에서 새롭게 유방암을 진단받은 환자 중 수술 전 1년 이내에 본 기관에서 선별 유방촬영술을 받은 환자는 49명 (39-76세, 평균 53.2세)이었으며, 당시 정상 (negative) 또는 양성 (benign)으로 판독된 33명의 환자 (40-76세, 평균 54.2세)가 이 연구에 포함되었다. 모든 선별 유방촬영술은 유방암 진단 당시의 유방촬영술보다 80-326일 이내에 시행되었다 (중양값, 231일).

모든 표준 디지털 유방촬영술 (full-field digital mammography) 영상은 상용화된 장비 (GE Senographe Essential; Hologic Selenia Dimensions)를 사용하여 획득되었고, 양측 유방에 대해 표준 상하위 (craniocaudal view, CC view)와 내외사위 (mediolateral oblique view, MLO view) 영상 촬영을 시행하였다. 유방촬영술에 대한 판독은 미국방사선의학회에서 정한 유방영상판독 및 데이터체계 (Breast Imaging Reporting and Data Systems; BI-RADS)의 권고에 따라 범주를 0에서 5까지 6단계로 분류하였고, 범주 0의 경우 재검 후 최종 범주로 재분류하였다. 이후 최종 범주 3, 4, 5를 양성으로 정의하

였다. 간격암은 새롭게 암을 진단받은 환자 중에서 1년 이내에 시행한 이전 선별 유방촬영술이 음성 (최종 범주 1 또는 2) 평가를 받은 경우로 규정하였다. 각각 1년, 13년의 경험이 있는 두 명의 유방 영상 전문의가 함께 의료영상저장전송시스템 (Picture Archiving and Communication System, PACS)을 이용하여 수술 직전 유방촬영술이나 침위치결정술에서 파악된 병변의 위치를 토대로 수술 이전 선별 유방촬영술에서 간격암의 유무 및 위치를 재평가하였다. 재평가시 이전 선별 유방촬영술에서 소환 (recall)이 필요했던 (actionable) 경우 놓친암으로, 후향적인 분석에도 불구하고 소환이 불가능했던 경우 진성 간격암으로 분류하였다.

합성곱 신경망 (convolutional neural network, CNN)에 기반한 딥러닝 알고리즘을 이용한 AI-CAD 프로그램은 Lunit INSIGHT MMG Ver.1.1.1.0 (Lunit, Seoul, Korea)을 이용하였다. AI-CAD 프로그램은 4개의 영상에 대해 각 유방 별로 비정상 점수를 매기고 역치인 10%를 기준으로 그 이상의 점수와 위치를 히트맵으로 보여주었다. AI-CAD 프로그램이 등고선으로 표시한 영역을 기준으로 해당 위치의 표식을 AI-CAD 검사 양성으로 인정하였고, 실제암의 종괴 혹은 석회화의 어느 부분이든 겹치는 부위에 히트맵이 그려진 경우 암 진단을 제대로 한 것으로 인정하였다. 실제 암이 아닌 다른 부위에 있는 모든 표식은 위양성으로 간주하였다. 병변의 위치는 진단 당시 유방촬영술 및 수술시 시행한 침위치결정술 확인 유방촬영술에서의 위치를 기준으로 판단하였다.

놓친암에 대한 AI-CAD 진양성 표시가 암 해당위치의 상하위 또는 내외사위 중에 적어도 하나가 있는 경우, 암을 찾은 것으로 인정하였다. 진성 간격암으로 분류된 사례에 대해 AI-CAD가 양성 표시를 한 경우도, 진단 당시 유방촬영술 및 수술시 침위치결정술 이후 확인 유방촬영술에서의 위치를 확인하여 진양성과 위양성 유무를 판단하였다. 이후, 선별 유방촬영술에서의 놓친암과 진성 간격암의 임상, 영상 및 병리학적 특징과 AI-CAD의 진단 결과에 대하여 두 명의 영상의학과 의사가 후향적으로 분석 기술하였다.

결 과

총 33예의 간격암을 가진 환자 중 15명은 만져지는 종괴 (8명), 유두 분비물 (4명), 유방통증 (3명)을 주소로 검진 주기 이전에 재내원하였다. 나머지 18명은 무증상이었으나, 선별 초음파 의심소견으로 재내원하였다. 전체 33예

Table 1. Clinical, Imaging and Histopathologic Characteristics of 33 Reassessed Interval Cancers[†]

| Characteristics | Missed cancer (n=12) | True interval cancer (n=21) |
|--|----------------------|-----------------------------|
| Age (range) | 55.8 ± 8.0 (40-66) | 53.3 ± 10.4 (41-76) |
| < 50 years | 2 (16.7%) | 9 (42.9%) |
| ≥ 50 years | 10 (83.3%) | 12 (57.1%) |
| Breast composition category | | |
| a (almost entirely fatty) | 0 (0.0%) | 0 (0.0%) |
| b (scattered fibroglandular density) | 2 (16.7%) | 3 (14.3%) |
| c (heterogeneously dense) | 6 (50.0%) | 11 (52.4%) |
| d (extremely dense) | 4 (33.3%) | 7 (33.3%) |
| Pathology | | |
| IDC | 6 (50.0%) | 12 (57.2%) |
| ILC | 3 (25.0%) | 1 (4.8%) |
| Mucinous carcinoma | 2 (16.7%) | 1 (4.8%) |
| DCIS | 1 (8.3%) | 6 (28.6%) |
| Metastasis to the breast | 0 (0.0%) | 1 (4.8%) |
| Node positive | 1 (8.3%) | 3 (14.3%) |
| Cancer type, grade and receptor status | | |
| Invasive Carcinoma, histologic grade | | |
| High | 3 (27.3%) | 2 (14.3%) |
| Intermediate | 6 (54.5%) | 9 (64.3%) |
| Low | 2 (18.2%) | 3 (21.4%) |
| DCIS, nuclear grade | | |
| High | 0 (0.0%) | 3 (50.0%) |
| Intermediate | 1 (100.0%) | 2 (33.3%) |
| Low | 0 (0.00%) | 1 (16.7%) |
| Estrogen receptor status | | |
| Positive | 7 (58.3%) | 18 (85.7%) |
| Negative | 5 (41.7%) | 2 (9.5%) |
| Unknown | 0 (0.0%) | 1 (4.8%) |
| Progesterone receptor status | | |
| Positive | 6 (50.0%) | 15 (71.4%) |
| Negative | 6 (50.0%) | 4 (19.1%) |
| Unknown | 0 (0.0%) | 2 (9.5%) |
| HER2 status | | |
| Positive | 3 (25.0%) | 2 (9.5%) |
| Negative | 9 (75.0%) | 17 (81.0%) |
| Unknown | 0 (0.0%) | 2 (9.5%) |

IDC = invasive ductal carcinoma, ILC = invasive lobular carcinoma, DCIS = ductal carcinoma in situ, HER2 = human epidermal growth factor receptor type

[†]Data is reported as mean±standard deviation for age and number(%) for categorical variable

Table 2. Result of AI-CAD Application for 33 Reassessed Interval Cancers

| | Missed cancer (n=12) | True interval cancer (n=21) |
|---------------------------------|----------------------|-----------------------------|
| AI-CAD detection (+) | 7 | 2 |
| Detected only on CC-view | 1 | 0 |
| Detected only on MLO-view | 3 | 1 |
| Detected both CC- and MLO-views | 3 | 1 |
| AI-CAD detection (-) | 5 | 19 |

AI-CAD = Artificial intelligence-computer assisted diagnosis, CC = craniocaudal, MLO = mediolateral oblique

의 간격암 중, 이전 유방촬영술을 재평가시 소환이 필요했던 경우가 12예 (36.4%)로 놓친암에 해당하였으며 소환이 불가능했다고 판단한 경우가 21예 (63.6%)로 진성 간격암에 해당하였다. 전체 33예 중 유방촬영술상 섬유유선조직 밀도는 지방형 유방 0예 (0.0%), 산재된 섬유유선 조직형 5예 (15.2%), 비균질 치밀유방 17예 (51.5%), 고도 치밀유방 11예 (33.3%)로 대부분이 치밀 유방이었다. 놓친암 중에서 83.3% (10/12)가, 진성 간격암 중에서 85.7% (18/21)가 치밀 유방이었다. 최종 병리조직결과는 침윤암 25예 (75.8%), 관상피내암 7예 (21.2%), 그리고 전이암 1예 (3.0%)였다. 놓친암과 진성간격암의 임상 특성, 영상 및 병리학적 특징은 Table 1에 기술하였다.

AI-CAD 프로그램의 전체적인 성적은 Table 2에 기술하였다. AI-CAD 프로그램은 전체 33예 중 총 9예 (27.3%, [9/33])의 암을 추가적으로 찾았고 그 중 놓친암이 7예 (58.3%, [7/12]), 진성 간격암이 2예 (9.5%, [2/21]) (그림 1)였다. 병변의 형태는 종괴가 6예, 석회화가 1예, 구조왜곡 또는 비대칭이 2예였다 (Table 3). AI-CAD 프로그램은 상하위와 내외사위 영상에서 총 36개 (19예)의 표시를 하였는데 그 중 11개는 종괴, 7개는 석회화, 18개는 구조왜곡 또는 비대칭으로 보였다. 총 36개의 표시중 실제암에 표시된 것은 13개 (36.1%, [13/36])였다. AI-CAD가 성공적으로 검출한 총 9예의 병리 조직의 평균 크기는 2.1 cm (범위, 0.8 - 4.9 cm)였다. 한편, 영상의학과 의사가 후향적으로 재평가시 소환이 가능하다고 판단했던 놓친암 12예의 평균 크기는 2.2 cm (범위, 0.3 - 4.9 cm)였다.

AI-CAD 프로그램이 암을 찾지 못한 사례는 총 24예로, 놓친 암이 5예 (그림 2), 소환이 불가능했던 진성 간격암이 19예 (그림 3)였다. 영상의학과 의사와 AI-CAD

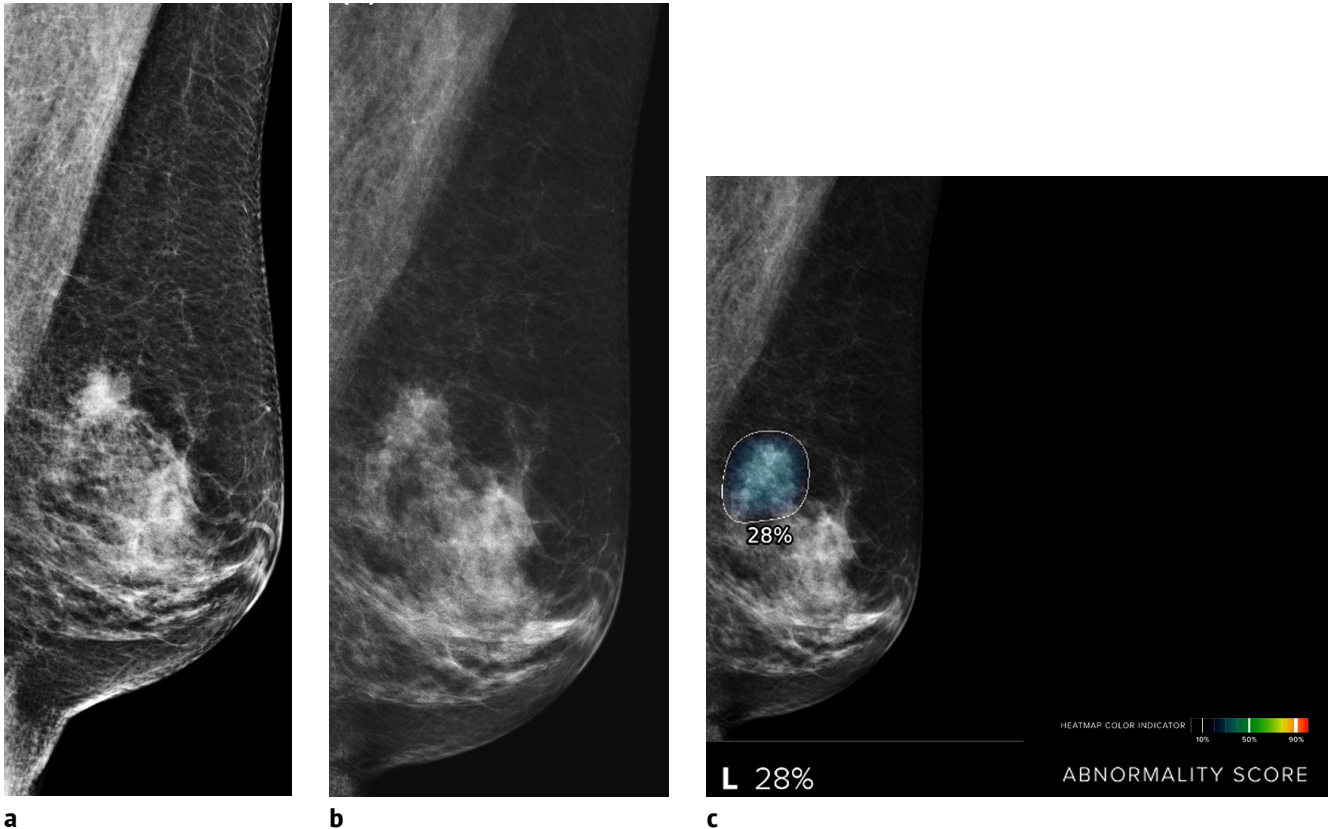


그림 1. 영상의학과 의사는 발견하지 못하였으나 AI-CAD가 발견한 진성 간격암
 (a) 51세 여성에서 BRCA1 변이로 고위험군 정밀 검사를 위해 시행한 유방촬영술. 좌측 내외사위 영상에서 상외측에 종괴 음영이 관찰되었다. (b) 6개월전에 시행된 선별 유방촬영술 상에서는 해당 위치의 음영은 섬유유선조직의 겹침으로 판단하여 진성 간격암으로 분류되었다. (c) 그러나, AI-CAD 프로그램은 해당 부분에 대한 암예측도를 28%로 양성으로 판단하였다. 조직검사 및 수술 결과 침윤성 유관암으로 진단되었다.

가 모두 음성으로 판단한 19예의 암은 각각 침윤암 13예 (68.4%), 관상피내암 5예 (26.3%), 그리고 전이암 1예 (5.3%)였다. 병리 조직의 평균 크기는 1.8 cm (범위, 0.1 - 7.0 cm)로 다소 작은 크기였다. 영상의학과 의사와 AI-CAD가 모두 음성으로 분류한 19예 중에서 치밀 유방은 84.2% (16/19)였고, 선별 유방촬영술뿐 아니라 진단 당시 유방촬영술에서도 암 위치를 특정하기 어려운 경우가 47.4% (9/19)였다.

AI-CAD 프로그램은 총 23개의 위양성을 표시하였고, 그 중 종괴가 2개, 석회화가 6개, 비대칭이 15개였다 (Table 3). 상하위와 내외사위를 각각 하나의 개별 영상으로 하였을 때, 한 영상 당 위양성 표시 (False Positive Per Image, FPPI)는 0.17개 (23/132) (범위, 0-4개 표시; 중앙값, 0개 표시)였다. 위양성 표시는 총 10예에서 23개로, 그 중 3예 (30.0%)의 농친암에서 7개 (30.4%), 7예 (70.0%)의 진성 간격암에서 16개 (69.6%)였다 (Table 3).

고 찰

본 연구의 결과에 따르면, 선별 유방촬영술 결과 정상판정 후 1년 내에 유방암이 진단된 33명의 여성에서, AI-CAD를 통해 총 9개의 암을 정확하게 진단해 낼 수 있었다. 이들 중 대부분인 7예는 농친암으로 분류할 수 있는 유방촬영에서 검출되었고, 2예는 다시 봐도 진단하기 어려운 진성 간격암에서 검출되었다.

간격암은 선별 유방촬영술 결과 정상판정 후 다음 선별 유방촬영술 시기 전에 발견되는 암을 말한다. 간격암은 후향적으로 다시 봐도 이상이 없어 보이는 진성 간격암, 관찰자 실수로 간과되거나 보이기 는 하나 비특이적이어서 진단이 어려운 농친 암, 유방촬영술에서 안 보이는 암 등의 이유로 발생하게 된다. 이들 중 농친암은 치밀 유방에 의한 가려짐 혹은 판독자의 피로도로 인한 인식 저하, 판독자의 경험 부족에 의한 해석 오류 등에 의해 발생

Table 3. Mammographic Features of 36 AI-CAD Markers among 33 Reassessed Interval Cancers

| | Missed cancer | | | True interval cancer | | |
|--------------------------------------|------------------------------------|----------------------------|----------------|------------------------------------|----------------------------|----------------|
| | True positive | | False positive | True positive | | False positive |
| | Detected on either CC- or MLO-view | Detected on both two views | | Detected on either CC- or MLO-view | Detected on both two views | |
| Breast composition category | No. of cases | | | | | |
| a (almost entirely fatty) | 0 | 0 | 0 | 0 | 0 | 0 |
| b (scattered fibroglandular density) | 2 | 0 | 1 | 0 | 0 | 1 |
| c (heterogeneously dense) | 1 | 1 | 0 | 1 | 1 | 4 |
| d (extremely dense) | 1 | 2 | 2 | 0 | 0 | 2 |
| Total | 4 | 3 | 3 | 1 | 1 | 7 |
| Dominant lesion feature | No. of marks (No. of cancer cases) | | | | | |
| Mass | 2 (2) | 4 (2) | 0 | 1 (1) | 2 (1) | 2 |
| Calcification | 1 (1) | - | 3 | - | - | 3 |
| Distortion or asymmetry | 1 (1) | 2 (1) | 4 | - | - | 11 |
| Total | 4 (4) | 6 (3) | 7 | 1 (1) | 2 (1) | 16 |

Table shows cancers detected by AI-CAD in which the abnormality score is greater than or equal to the predefined threshold of 0.1
 AI-CAD = Artificial intelligence-computer assisted diagnosis, CC = craniocaudal, MLO = mediolateral oblique, No. = number

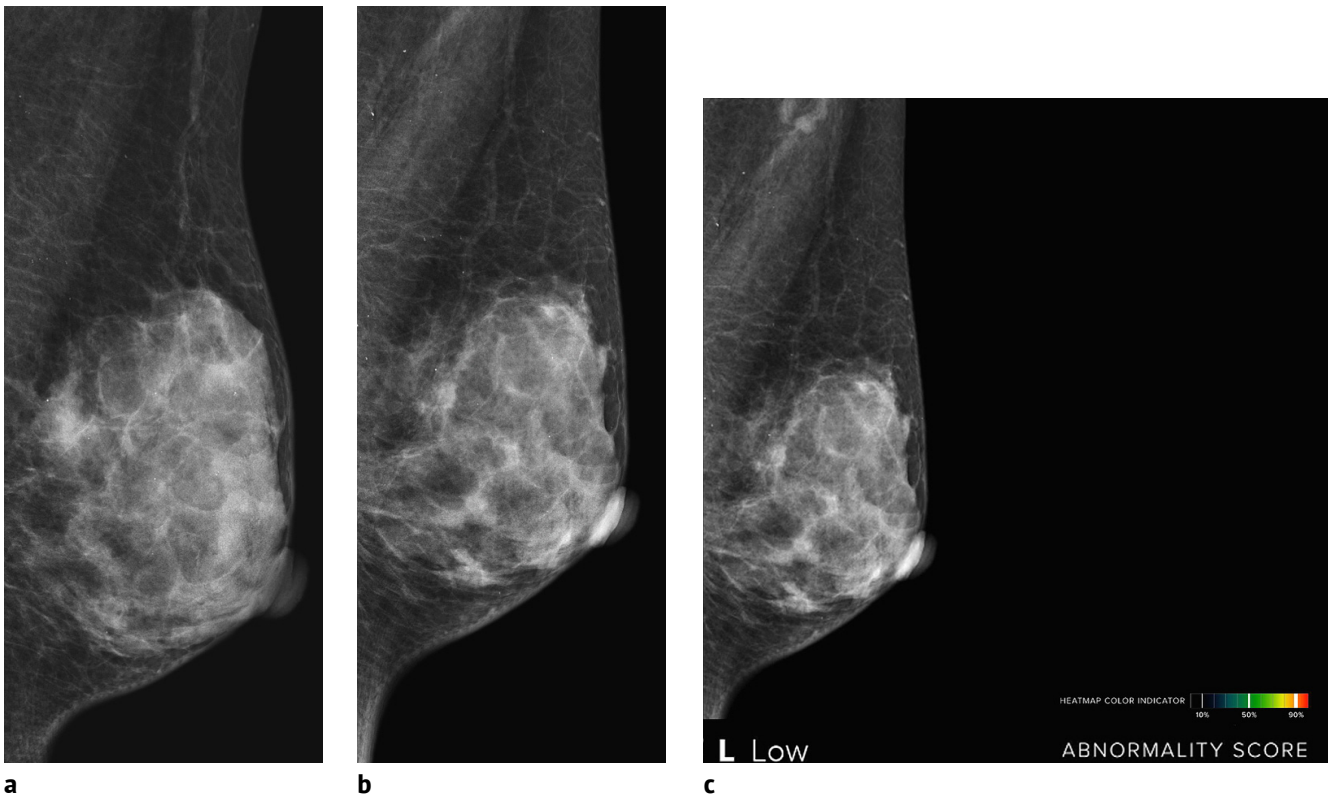
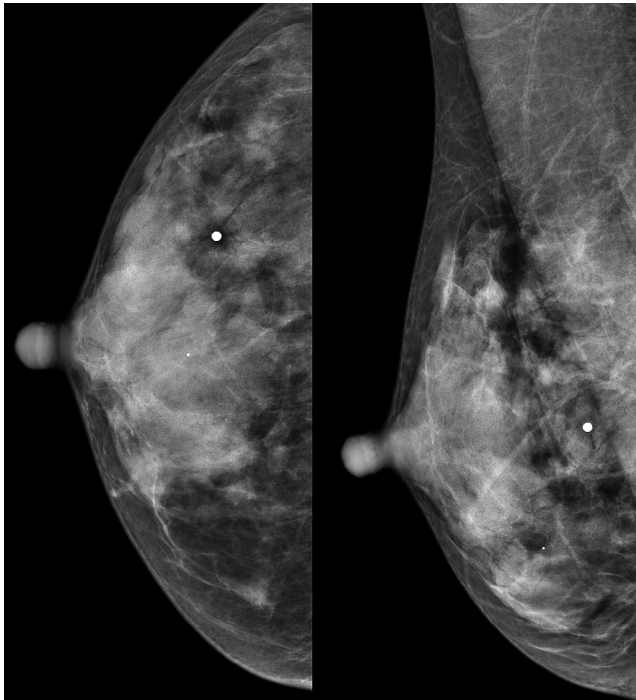
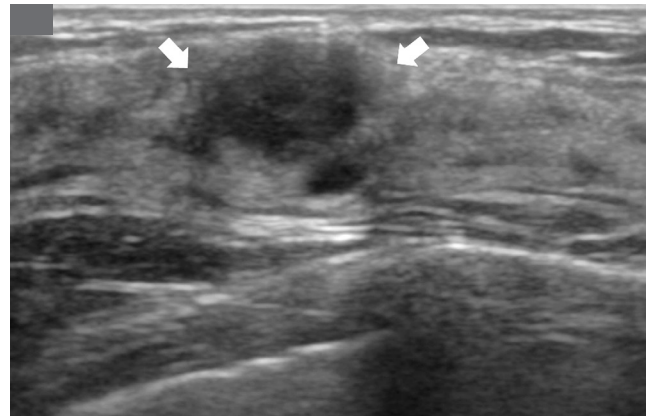


그림 2. AI-CAD가 발견하지 못한 농친암

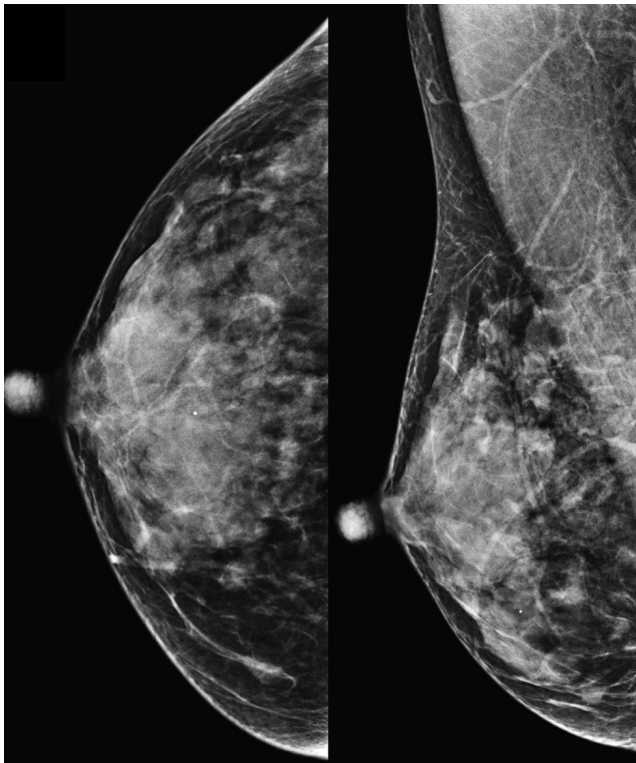
(a) 만저지는 종괴를 주소로 내원한 39세 여성의 유방촬영술. 좌측 내외사위 유방촬영술상 상내측의 종괴가 관찰되었다. (b) 9개월 전 무증상 검진으로 시행한 좌측 내외사위 선별 유방촬영술상 상내측 종괴가 덜 뚜렷하나 재평가시 보이는 상태로 농친암으로 분류되었다. (c) AI-CAD 프로그램은 해당 영상을 음성으로 판단하였다. 조직검사 및 수술 결과 침윤성 유관암으로 진단되었다.



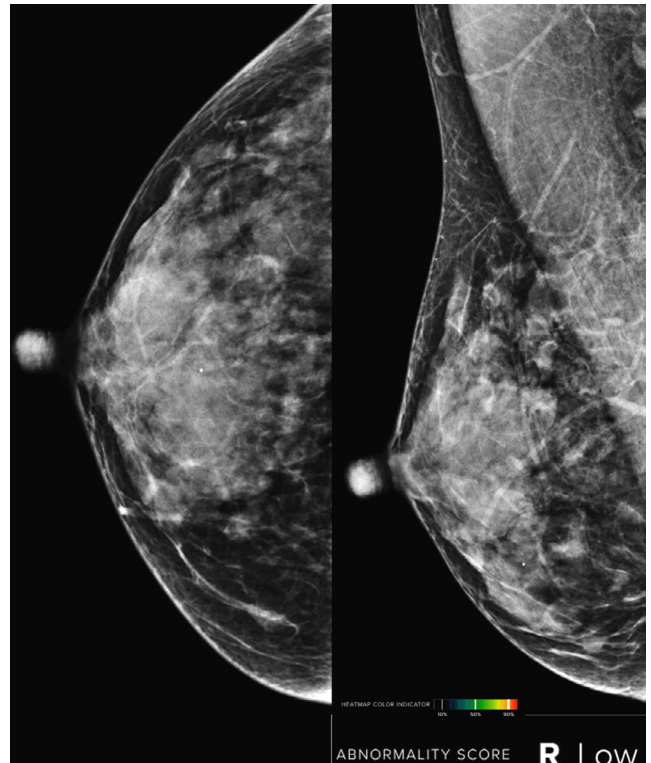
a



b



c



d

그림 3. AI-CAD와 영상의학과 의사가 모두 발견하지 못한 진성 간격암
 (a) 만져지는 우하외측 종괴를 주소로 내원한 50세 여성의 유방촬영술. 치밀유방 소견 외에 특이 소견이 없었다. (b) 같은 날 시행한 초음파 검사에서 해당 위치에 의심스러운 모양의 종괴가 발견되었다. (c) 6개월 전에 시행한 선별 유방촬영술에서는 특이소견을 보이지 않아 진성 간격암으로 평가되었다. (d) AI-CAD 프로그램 적용 결과 음성으로 평가하였다. 조직검사 및 수술 결과 침윤성 유관암으로 진단되었다.

한다 (7). 아시아 여성에서는 치밀 유방의 비율이 높아 간격암의 발생률도 높다 (8). 본 연구의 놓친암 중 83.3% (10/12)가 치밀 유방이었고 이는 치밀 유방으로 인한 병변 인식의 어려움과 판독자의 피로도로 인한 주의력 저하 등이 암의 진단에 영향을 미쳤음을 알 수 있다. 본 연구에서는 상기와 같이 다양한 이유로 당시 유방촬영술에서 놓친암에 대해서 AI-CAD가 약 60%의 암을 찾아냈고, 치밀 유방에서 놓친 암 중에서도 50.0% (5/10)를 찾아냈다. 놓친암에서 75.0% (9/12)가 종괴였는데 이는 치밀 유방에서 종괴로 발현하는 연조직 병변 (soft tissue lesion) 발견의 어려움을 시사한다. 이에 AI-CAD를 이용할 때 유방 실질에서 연조직 병변을 구별해내는데 도움을 받을 수 있고 이를 통한 인식의 제고를 통하여 판독 오류를 줄이고, 판독 시간 단축도 가능할 것으로 사료된다 (2).

한편, 영상의학과 의사가 재평가한 후에도 소환이 불가능했던 총 21예의 진성 간격암 중 2예에서 AI-CAD가 유방암을 정확하게 검출하였고, 나머지 19예는 검출하지 못했다. 진성 간격암의 경우, AI-CAD의 도움을 받아 영상의 해석에 일부 도움을 받을 수 있으나 위양성 중 약 70.0%가 진성 간격암에서 발생하였음을 고려할 때 영상의학과 의사가 다시 봐도 알기 어려운 진성 간격암에서 AI-CAD의 효용성은 아직 불확실하며, 유방촬영술 외에 추가적인 보조 진단도구의 필요성이 제기된다.

인공지능을 이용한 많은 연구들이 의료 영상에 적용되고 있으며, 유방촬영술도 그 중 큰 분획을 차지한다. 많은 선별검사 작업량 (workload)으로 인해 가중되는 영상의학과 의사의 피로도로 인해 놓친암이 발생함을 고려할 때, 유방영상에 숙련된 영상의학과 의사의 경우 인식의 측면에서 보조적으로 AI-CAD의 도움을 받아 시간을 절약할 수 있고, 비숙련 판독자의 경우 해석의 측면에서 AI-CAD가 제시하는 암 예측도를 지표 삼아 위음성을 줄일 수 있을 것으로 기대한다.

본 연구의 제한점은 다음과 같다. 첫째, 연구에 포함된 대상수가 적다. AI-CAD의 성능에 대한 보다 객관적이고 정확한 평가를 위해서는 표본의 수가 좀 더 많아야 하고, 정상 대조군을 통한 리더스터리가 필요할 것으로 사료된다. 둘째, 간격암 발견에 있어 증상뿐만 아니라, 다른 검진을 통해 발견된 증례가 포함되어 연구 대상이 이질적인 측면이 있다. 셋째, 소환 필요여부에 대한 결정을 두 명의 영

상의학과 의사가 합의하에 평가했다는 점이다. 다양한 경험을 가진 영상의학과 의사가 더 많이 참여했으면 보다 정확한 판단이 가능했을 것으로 생각된다.

결론적으로, AI-CAD는 선별 유방촬영술에서 발생하는 간격암을 줄이는데 AI-CAD가 도움이 되며, 특히 판독자의 실수나 해석 오류 등에 의한 놓친암의 발견에 좋은 진단능을 보일 것으로 생각된다. 향후 영상의학과 의사의 유방암 진단에 있어 인식 및 해석에 효과적인 도구로 사용될 것으로 전망된다.

참 고 문 헌

1. Myers ER, Moorman P, Gierisch JM, et al. Benefits and harms of breast cancer screening: a systematic review. *JAMA* 2015;314:1615-1634
2. Kim HE, Kim HH, Han BK, 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. Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology* 2009;253:641-651
4. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 2007;356:1399-1409
5. 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;175:1828-1837
6. Rodríguez-Ruiz A, Krupinski E, Mordang J-J, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. *Radiology* 2019;290:305-314
7. Majid AS, de Paredes ES, Doherty RD, Sharma NR, Salvador XJR. Missed breast carcinoma: pitfalls and pearls. *Radiographics* 2003;23:881-895
8. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;356:227-236

Retrospective Analysis for Interval Cancer in Screening Mammography Applied for Artificial Intelligence Based Computer Aided Diagnosis

Yeon Soo Kim, MD, Jung Min Chang, MD, PhD

*Department of Radiology, Seoul National University College of Medicine,
Seoul National University Hospital, Seoul, Korea*

Purpose: The purpose of this article is to retrospectively evaluate the diagnostic performance of artificial intelligence-computer aided diagnosis (AI-CAD) for interval cancers in screening mammography.

Materials and Methods: From January 2016 to December 2019, thirty-three breasts of 33 women (age range 40–76, mean age 54.2 years) who were newly diagnosed with breast cancer and underwent preoperative mammography, whose screening mammography was negative within one year, were included. Two dedicated breast radiologists reassessed the previous screening mammography in consensus, and thirty-three interval cancers were classified into missed and true interval cancers. The reassessment results were compared with those of AI-CAD.

Results: In reassessment, thirty-three interval cancers were classified into 12 missed cancers (36.4%) determined to be actionable for recall and 21 true interval cancers (63.6%) determined to be insufficient for recall. Of the 36 AI-CAD marks, thirteen (36.1%) were marked at the true positive cancer. Of the 33 interval cancer, AI-CAD had correctly depicted 9 interval cancers (27.3%, [9/33]) including 7 missed cancers (58.3%. [7/12]).

Conclusion: AI-CAD had correctly marked a substantial number of missed cancer determined to be actionable. Application of AI-CAD for screening mammography is expected to reduce missed cancers caused by the interpretation error of radiologists.

Index words: Breast; Mammography; Missed cancer; Artificial intelligence

Corresponding author: Jung Min Chang, M.D., Ph.D.