

자동유방초음파와 인공지능: 판독자 경험에 따른 Multiview Convolutional Neural Network의 진단 성능

최서윤, 최은정

전북대학교 의과대학 영상의학과, 전북대학교병원 의생명연구원 임상의학연구소

자동 유방 초음파 (Automated Breast Ultrasound, ABUS)는 치밀 유방에서의 보조 선별검사로 활용되고 있으나, 방대한 3차원 데이터로 인해 판독 시간 증가와 판독자 간 성능 차이라는 한계를 가진다. 최근 합성곱 신경망 (convolutional neural network, CNN)를 기반으로 인공지능(Artificial Intelligence, AI), 특히 횡단면 (transverse view)와 관상면 (coronal view)를 동시에 활용하는 다중 시점 합성곱 신경망 (multiview convolutional neural network, multiview CNN) 전략이 도입되면서 진단 성능 향상이 보고되고 있다. 본 종설은 ABUS 기반 multiview CNN의 진단 성능을 종합적으로 정리하고, 판독자 경험 및 영상 특성에 따른 보조 효과의 차이를 분석함으로써 국내 검진 환경에서의 임상적 적용 방향을 제시하고자 한다.

Index words: Breast; Automated breast ultrasound; Multiview convolutional neural network; Ultrasound features

서론

유방암은 전 세계 여성에서 가장 흔한 암이며, 암 사망 원인 중 두 번째로 높은 비율을 차지한다 (1). 유방암의 조기 진단은 사망률 감소에 중요한 역할을 한다. 유방촬영술은 조기 유방암 진단을 위한 기본 영상 검사이지만, 진단율을 더욱 향상시키기 위해서는 추가적인 영상 검사가 필요하다 (2, 3).

자동 유방 초음파 (Automated Breast Ultrasound, ABUS)는 수기 초음파 (Hand-held Ultrasound, HHUS)에 비해 전 유방을 일정하고 재현 가능하며 검사자 의존성이 낮은 방식으로 평가할 수 있는 검사법이다. 최근 ABUS는 치밀 유방 여성에서 유방촬영술과 병행할 경우 HHUS와 유사한 진단 성능을 보이는 보조 선별검사로 부각되고 있다 (4, 5). 그러나 숙련된 영상의학과 전문의라 하더라도 실제 임상에서 수백 장에 이르는 ABUS 영상을 판독하는 데에는 많은 시간이 소요된다 (6). 이에 따라 여러 연구자들이 ABUS 영상에 영상 처리 기술과 딥러닝(deep learning)을 적용해 왔다 (4, 7).

합성곱 신경망 (convolutional neural network, CNN)는 대표적인 딥러닝 구조 중 하나로, 영상 분류 작업에서 유용성이 입증되어 왔다. 유방암 분류를 위한 컴퓨터 보조 진단 시스템에 CNN을 통합하려는 시도도 다수 이루어졌다 (8-10). 특히 multiview CNN은 여러 초음파 스캔으로부터 병변의 문맥 정보를 동시에 획득할 수 있어 ABUS

Corresponding author: Eun Jung Choi
Department of Radiology and Research Institute of Clinical Medicine of Jeonbuk National University-Biomedical Research Institute of Jeonbuk National University Hospital, Jeonbuk National University Medical School, 20 Geonji-ro, Deokjin-gu, Jeonju City, Jeollabuk-Do 54907, Republic of Korea
Tel: +82-63-250-1150 Fax: +82-63-272-0481
e-mail: cejcej80@hanmail.net

에 적용되었으며, 유방암 분류 성능 향상에 기여한 바 있다 (9). 임상 현장에서 multiview CNN을 신뢰성 있게 적용하기 위해서는, 영상학과 의사가 유방 병변을 감별하는 과정에서 CNN 성능에 영향을 미치는 다양한 요인을 규명하는 것이 중요하다. 기존 연구에서는 병변 크기가 딥러닝 시스템 및 CNN의 ABUS 성능에 영향을 미칠 수 있다고 보고하였으나, 이는 병변 검출(detection)에 관한 요인이었으며 병변 분류(classification)에 대한 분석은 아니었다 (10). 한편, CNN의 병변 분류 성능은 판독자의 경험 수준과 다양한 초음파 영상 특징에 의해 영향을 받을 수 있다고 보고된 바 있다. (11).

이 종설의 목적은 ABUS에서 multiview CNN의 진단 성능을 종합적으로 고찰하고, 판독자 경험 수준과 병변 유형, 배경에코, 유선조직 구성, 병변 크기와 같은 영상 특징이 AI 보조 효과에 어떠한 영향을 미치는지를 정리함으로써, 실제 임상 검진 환경에서 multiview CNN을 보다 효과적으로 적용할 수 있는 근거를 제시하는 데 있다.

ABUS에서 딥러닝 기반 AI의 발전

기존 초음파 컴퓨터 보조 진단 (computer-aided detection, CAD) 시스템은 Histogram of Oriented Gradients, Principal Component Analysis 등의 수작업 기반 특징 추출 기법과 Support Vector Machine 분류기를 활용하였다 (12). 그러나 이러한 접근은 특징 설계 과정에 연구자의 주관이 개입될 수 있으며, 다양한 영상 조건에 대한 일반화 성능이 제한적이라는 한계를 지녔다. 또한 병변의 복잡한 질감이나 미세한 구조적 차이를 충분히 반영하기 어려웠다.

이에 비해 딥러닝 기반 CNN은 영상으로부터 특징을 직접 학습함으로써 수작업 기반 설계의 한계를 극복할 수 있다. 전이학습을 통해 ImageNet과 같은 대규모 데이터셋에서 사전 학습된 네트워크를 의료 영상에 적용하면, 제한된 데이터 환경에서도 비교적 안정적인 성능을 확보할 수 있다 (13). 특히 Inception-v3와 같은 구조는 다양한 크기의 convolution kernel을 병렬적으로 적용하여 병변의 다중 스케일 특징을 효과적으로 추출할 수 있으며, 질감과 구조 정보가 혼재된 초음파 영상에 적합한 특성을 보인다.

ABUS는 3차원 볼륨 데이터로 구성되며 다평면 재구성 (multi-planar reconstruction) 기술을 사용하여 세 가지 축을 동시에 보여준다. 특히, 관상면 (coronal view)은 견인 현상 (retraction phenomenon)과 같은 구조적 단서를 포함하고 있어 병변의 공간적 특성 평가에 중요한 정보를 제

공한다. 이러한 다면적 정보는 단일 단면 기반 분석보다 풍부한 문맥 정보를 제공하므로, 다중 시점 학습 전략이 자연스럽게 적용될 수 있다. Multiview CNN은 서로 다른 평면에서 획득된 병변 정보를 병렬적으로 학습한 뒤 이를 통합하여 최종 분류를 수행함으로써 병변의 형태, 경계, 주변 조직과의 관계 등 공간적 문맥을 동시에 반영할 수 있다는 장점을 지닌다. 그렇기 때문에 기존 단일 시점 기반 모델에 비해 높은 성능이 보고되었으며, 특히 작은 병변이나 비정형 병변에서 성능 향상이 두드러진다. 따라서 multiview CNN은 ABUS의 3차원 구조적 특성과 가장 잘 부합하는 딥러닝 전략으로 이해할 수 있다. 이러한 다중 시점 학습 전략의 임상적 의미를 보다 명확히 이해하기 위해, 먼저 초음파 및 ABUS 기반 AI의 진단 보조 효과에 대한 기존 연구 결과를 살펴볼 필요가 있다.

이전 연구에서 보고된 HHUS 및 ABUS 기반 AI의 진단 성능

유방 초음파 영역에서 AI의 진단 보조 효과는 여러 상용 제품과 다기관 연구를 통해 점차 입증되고 있다. HHUS에서 AI를 병합한 연구에서는 AI 보조 시 특이도와 정확도가 유의하게 향상되는 경향이 반복적으로 보고되었다. 이전 연구에 의하면 경험이 있는 판독자에서도 민감도 변화 없이 특이도는 약 72.8%에서 93.1%로 증가하였고 ($p < 0.001$), 정확도 또한 77.9%에서 90.9%로 향상되었다. 이는 AI가 위양성을 감소시키는 방향으로 기여함을 시사한다 (14).

이러한 경향은 HHUS를 이용한 다른 다기관 연구에서도 유사하게 관찰되었다. HHUS에서 CadAI-B (CadAI Inc., Seoul, Republic of Korea)를 이용한 다기관 연구에서도 유사한 경향이 확인되었다 (15). 42명의 다양한 수준의 판독자를 대상으로 한 분석에서, 경험이 적은 판독자 그룹($n=38$)은 baseline 기준 상태의 수신자 조작 특성 곡선 하 면적 (area under the receiver operating characteristic curve, AUC)가 0.661에서 AI 적용 후 0.759로 증가하였고 ($p=0.029$), 민감도 역시 71.1%에서 89.6%로 향상되었다 ($p=0.033$). 반면 경험이 많은 판독자($n=4$)에서는 AUC가 0.727에서 0.795로 증가하였으나 통계적으로 유의하지 않았다. 이는 HHUS에서 AI의 보조 효과가 판독자 경험 수준에 따라 다르게 나타날 수 있음을 보여준다.

한편, 이러한 인공지능 보조 효과는 ABUS환경에서도 보고되고 있다. ABUS에서 전용 AI 시스템 (MONCAD ABS DX; MONCAD Co., Ltd., Seoul, Republic of Korea)를

이용한 연구에서도 AI 활용 방식에 따른 성능 차이가 보고되었다. 전문의 단독 판독 시 AUC는 0.87, 민감도 77.6%, 특이도 88.0%였으며, AI를 참고하여 전문의가 최종 판단을 내리는 방식에서는 AUC가 0.919로 증가하고 특이도도 92.3%로 향상되었다. 반면 AI 단독 판독에서는 AUC가 오히려 감소하였고 특이도 또한 낮아지는 경향을 보였다. AI 결과에 따라 등급을 상향하거나 하향하는 방식에 따라 민감도와 특이도의 균형이 달라지는 양상도 관찰되었다. 이러한 결과는 AI가 독립적 진단 도구로 사용될 때보다, 전문의의 판단을 보조하는 참고 도구로 활용될 때 보다 안정적이고 균형 잡힌 성능을 보일 수 있음을 시사한다 (16).

이와 더불어 최근에는 ABUS 영상 자체를 기반으로 한 딥러닝 모델 연구도 활발히 보고되고 있다. 최근 2020년부터 2024년까지 발표된 약 20여 편의 연구를 종합하면, ABUS 기반 딥러닝 분류 모델의 AUC는 대체로 0.85에서 0.99 범위로 보고되고 있으며, 이는 우수한 진단 판별 능력을 의미한다. 민감도는 85-96%, 특이도는 70-95% 수준으로 보고되었다. Detection 모델의 경우 90% 이상의 민감도를 유지하면서 volume 당 2개 이하의 위양성을 달성한 연구도 보고되었으며, DenseNet, DeepLab-v3, Inception-v3 기반 구조, MVSA-Net 등 다양한 딥러닝 아키텍처가 적용되었다. 이는 ABUS 기반 AI가 기술적으로 높은 수준의 진단 성능에 도달했음을 보여준다 (17).

기존 AI 연구의 한계와 고려 사항

이처럼 기존 연구들은 초음파 및 ABUS 영역에서 AI의 전반적인 진단 성능 향상을 일관되게 보고하고 있으나, 몇 가지 한계도 존재한다. 첫째, 대부분의 연구가 전체 환자군을 대상으로 평균적인 성능 지표 (AUC, 민감도, 특이도)를 제시하는 데 초점을 맞추어, 병변 유형 (종괴 및 비종괴 병변)이나 병변의 크기, 그리고 치밀 유방, 후방 음영, 주변 조직과의 대비 저하 등과 같은 복잡한 영상 조건에서의 세부 성능 차이를 충분히 분석하지는 못하였다 (9,14). 둘째, 판독자 경험 수준에 따른 보조 효과의 차이는 일부 연구에서 제시되었으나, 동일한 조건에서 성능 향상 정도를 체계적으로 비교하고 각 집단에서 AI의 기여도를 진단 지표별로 종합적으로 해석한 연구는 제한적이었다 (9,11). 셋째, 배경에코 (background echotexture), 선조직성분 (glandular tissue component, GTC), 병변 크기와 같은 영상 특성이 AI 성능에 미치는 영향은 충분히 평가되지 않았다.

그 외에 기술적·데이터 기반 한계 또한 존재한다. ABUS

는 3차원 볼륨 데이터를 제공하므로 높은 연산 자원을 요구하며, 서로 다른 단면 간 병변 정합(cross-slice alignment)이 필요하여 모델 학습과 추론 과정이 복잡하다. 또한 장비 제조사 간 영상 획득 방식과 재구성 알고리즘의 차이는 모델의 일반화 성능에 영향을 미칠 수 있다. 대규모 정밀 주석 데이터 확보가 어렵고, 공개된 다기관 ABUS 데이터셋이 부족하다는 점도 외부 검증과 재현성 확보에 제약으로 작용한다. 또한 임상 상황, 판독자, 병변 유형에 따른 AI 활용의 최적 조건에 대한 구체적인 가이드라인은 아직 명확히 제시되지 않았다. AI의 도입은 단순히 진단 정확도 향상이라는 기술적 성과에 그치는 것이 아니라, 판독 과정의 표준화와 일관성 확보, 교육적 보조 도구로서의 활용, 그리고 의료 자원의 효율적 배분과 같은 임상적·시스템적 측면까지 함께 고려되어야 한다. 특히 치밀 유방이 많은 국내 검진 환경에서는 판독자의 경험 격차를 줄이고 판독 편이를 최소화하는 것이 중요한 과제로 남아 있다.

따라서 기존 연구에서 제시된 평균 성능 지표는 ABUS 기반 AI의 전반적인 가능성을 보여주지만, 실제 임상 환경을 충분히 반영한다고 보기는 어렵다. 특히 병변의 유형, 배경에코, 선조직성분, 병변 크기와 같은 세부 영상 조건과 판독자 경험 수준에 따라 AI의 보조 효과가 달라질 가능성이 있음에도 불구하고, 이에 대한 체계적인 분석은 제한적이었다. 이러한 점은 ABUS 기반 AI의 임상적 의미를 보다 정밀하게 이해하기 위해 추가적인 고찰이 필요함을 시사한다.

판독자 및 영상 특성에 따른 Multiview CNN의 임상적 영향

1. 판독자 경험에 따른 Multiview CNN의 성능 변화

최근 ABUS 영상에서 multiview CNN의 보조 효과를 평가한 연구에서는 총 656개의 병변을 대상으로 6명의 판독자가 AI 보조 전후의 진단 성능을 비교하였다 (18). 전체 판독자의 평균 AUC는 AI 보조 전 0.84에서 보조 후 0.93으로 향상되었으며, 민감도와 특이도 또한 모두 증가하는 경향을 보였다. 이러한 결과는 multiview CNN이 단순히 특정 지표 하나를 개선하는 것이 아니라, 전반적인 진단 성능의 균형을 향상시킬 수 있음을 시사한다.

판독자 경험 수준에 따른 분석에서는 보다 의미 있는 차이가 관찰되었다. 경험이 적은 판독자에서는 baseline AUC가 0.79였으나, AI 보조 후 0.91로 크게 향상되었다.

특히 민감도의 증가가 두드러졌으며, 이는 비정형 병변이나 경계가 불명확한 병변에서 AI가 추가적인 판단 근거를 제공함으로써 위음성을 줄이는 데 기여했을 가능성을 시사한다. 경험이 적은 판독자의 경우 병변의 미세한 형태학적 특징이나 배경 조직과의 대비를 해석하는 데 어려움이 있을 수 있는데, multiview CNN은 서로 다른 평면에서의 정보를 통합하여 보다 안정적인 분류 결과를 제공함으로써 이러한 한계를 보완하는 역할을 한 것으로 해석된다.

반면 유방 전문의는 baseline AUC가 이미 0.93 수준으로 높았으며, AI 보조 후 0.96으로 유의한 향상이 관찰되었다. 향상 폭은 전문의가 아닌 이에 비해 상대적으로 작았으나, 이는 전문의의 높은 초기 성능을 고려할 때 의미 있는 결과로 볼 수 있다. 특히 복잡한 배경 조직이나 작은 병변과 같은 고난도 조건에서 AI가 추가적인 참고 정보를 제공함으로써 판독의 확신도를 높였을 가능성이 있다.

이러한 결과는 multiview CNN이 경험 격차를 줄이는 방향으로 작용할 수 있음을 보여준다. 즉, AI는 경험이 적은 판독자의 진단 정확도를 전문의 수준에 가깝게 향상시키는 보조 도구로 기능할 수 있으며, 동시에 숙련된 판독자에게도 선택적인 상황에서 추가적인 안정성을 제공할 수 있다. 더 나아가 관찰자 간 일치도 (interobserver agreement)의 향상은 AI가 판독자 간 변이를 감소시키고 진단 일관성을 높이는 데 기여할 수 있음을 시사한다. 따라서 multiview CNN은 단순한 정확도 향상 도구를 넘어, 판독자 경험 수준에 따른 성능 차이를 완화하고 판독 과정의 표준화를 지원하는 보조 기술로 이해할 수 있다.

2. 병변 유형에 따른 Multiview CNN의 성능 변화

병변 유형은 ABUS에서 판독 난이도를 결정하는 중요한 요소이다. 연구 결과에 따르면 종괴 병변 (mass lesion)에서는 multiview CNN 보조 후 AUC가 0.81에서 0.91로 향상되었으며, 전반적으로 안정적인 성능 개선이 관찰되었다. 종괴 병변은 비교적 경계가 명확하고 형태학적 특징이 뚜렷한 경우가 많아, baseline 성능 자체가 비교적 높은 경향을 보인다. 이러한 특성으로 인해 AI 보조 효과는 유의하지만, 향상 폭은 제한적일 수 있다.

반면 비종괴 병변(non-mass lesion)에서는 multiview CNN의 효과가 더욱 두드러졌다. 비종괴 병변의 baseline AUC는 0.56 수준으로 낮았으나, AI 보조 후 0.90까지 크게 증가하였다. 특히 유방 전문의가 아닌 이에서 민감도의 현저한 향상이 관찰되었는데, 이는 비종괴 병변이 구조 왜곡, 비대칭성, 미세한 음영 변화 등 간접적인 영상 소견

에 의존하여 진단되는 경우가 많기 때문으로 해석된다 (그림 1). 이러한 병변은 Breast Imaging Reporting and Data System (BI-RADS) 체계 내에서도 명확한 정의와 경험 축적이 필요한 영역으로, 숙련도에 따른 판독 차이가 크게 나타날 수 있다.

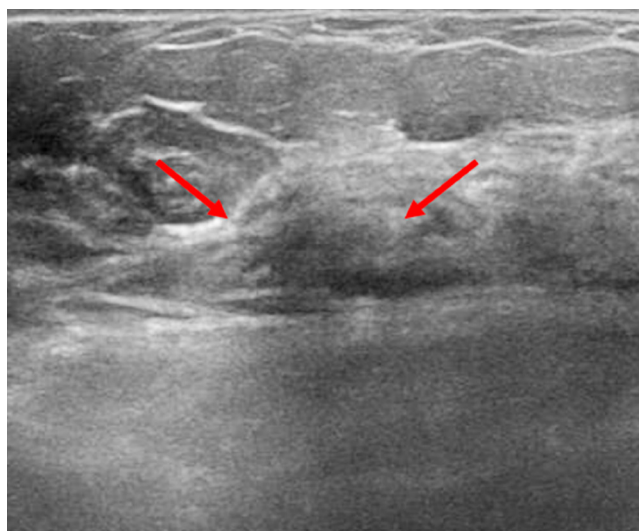
Multiview CNN은 종단면 및 관상면의 정보를 통합하여 병변의 형태뿐 아니라 주변 조직과의 공간적 관계까지 반영할 수 있다. 관상면에서 관찰되는 견인 현상이나 구조적 왜곡과 같은 특징은 단일 단면 영상에서는 충분히 평가하기 어려울 수 있다. 다중 시점 전략은 이러한 다면적 정보를 동시에 분석함으로써 비종괴 병변과 같이 해석이 어려운 병변에서 진단 안정성을 향상시킨 것으로 보인다.

이러한 결과는 AI의 보조 효과가 단순히 평균적인 성능 향상에 그치는 것이 아니라, 판독 난이도가 높은 병변 유형에서 특히 의미 있게 나타날 수 있음을 시사한다. 이는 실제 임상 환경에서 비종괴 병변과 같은 고난도 병변에서 AI가 보완적 판단 근거를 제공할 수 있다는 점에서 중요한 의미를 가진다.

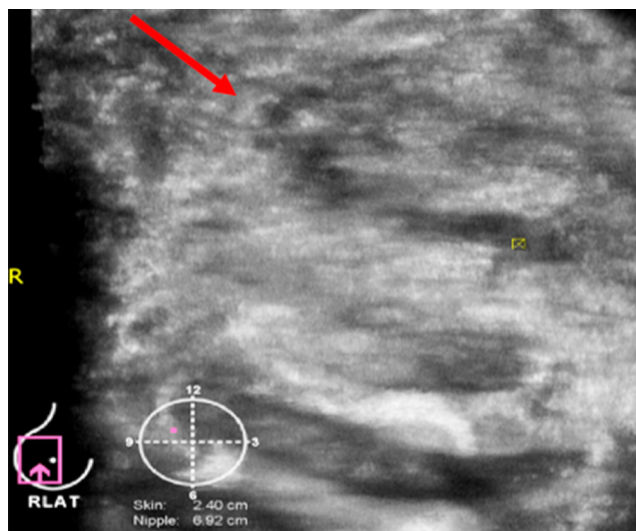
3. 배경에코에 따른 Multiview CNN의 성능 변화

배경에코는 ABUS 판독에서 병변 인식에 직접적인 영향을 미치는 중요한 영상 요소이다. 균질-지방형 (homogeneous-fat) 또는 균질-섬유유선형 배경에코 (homogeneous-fibroglandular background)에 비해 불균질형 배경에코 (heterogeneous background)는 구조적 복잡성이 증가하며, 병변과 정상 조직 간 대비가 감소하는 경향이 있다. 이로 인해 위양성 및 위음성 가능성이 모두 증가하고, 판독자의 주관적 해석 차이도 커질 수 있다. Multiview CNN은 모든 배경 유형에서 AUC를 유의하게 향상시켰으며, 특히 불균질형 배경 에코에서 개선 폭이 가장 컸다. 전문의에서도 불균질형 배경 에코 조건에서 AUC가 0.90에서 0.95로 유의하게 증가하였다. 이는 baseline 성능이 이미 높은 숙련된 판독자에서도 복잡한 배경 조직에서는 추가적인 보조 효과가 존재함을 의미한다.

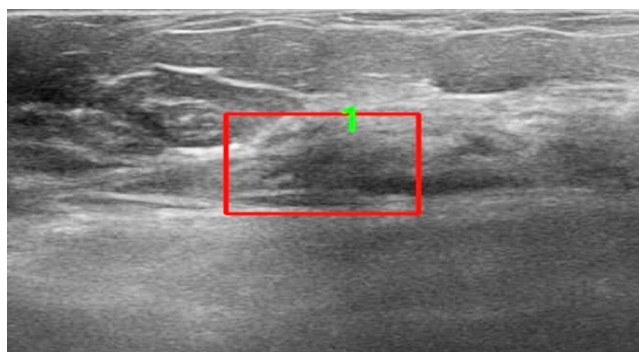
불균질형 배경 에코에서는 섬유유선조직과 지방조직이 혼재되어 있어 음영 패턴이 불균질하게 나타난다. 이러한 조건에서는 병변의 경계가 주변 조직과 혼동되기 쉽고, 미세한 형태학적 변화가 배경 신호에 의해 가려질 수 있다. Multiview CNN은 서로 다른 평면에서의 신호 패턴을 통합하여 병변과 정상 조직 간의 상대적 특징 차이를 학습하므로, 복잡한 배경 구조에서도 보다 안정적인 분류를 수행할 수 있는 것으로 보인다.



a



b



c

그림 1. 불균질형 배경에코와 중등도 선조직성분을 가진 52세 여성의 ABUS에서 관찰된 비종괴 병변으로 침습성 유방암으로 진단된 사례. (a) 횡단면에서 관찰된 병변 (화살표). (b) 관상면에서 관찰된 동일 병변 (화살표). (c). Multiview CNN이 검출한 병변 (네모 박스). Multiview CNN에 의한 악성 예측도는 0.999였다. Multiview CNN 보조 전과 후의 판독 결과는 다음과 같았다: 유방 영상 전문의가 아닌 이들은 BI-RADS category 3에서 5로 상향 판독하였다. 유방 영상 전문의들은 BI-RADS category 4에서 5로 판독하였다.

특히 국내 여성의 경우 치밀 유방 비율이 높다는 점을 고려하면, 불균질형 배경 에코에서의 성능 향상은 임상적으로 더욱 중요하다. 이는 AI가 실제 검진 환경에서 자주 마주하는 복잡한 영상 조건에서 판독 보조 도구로 활용될 수 있음을 시사한다.

또한 불균질형 배경 에코에 따른 성능 분석은 단순히 AI의 평균적 정확도를 평가하는 것보다, 임상 상황별 적용 전략을 수립하는 데 더 실질적인 정보를 제공한다. 예를 들어, 불균질형 배경 에코에서 AI 보조를 적극적으로 활용하는 전략은 판독 변이를 줄이고 위음성 위험을 감소시키는 데 기여할 수 있다.

4. 선조직성분(Glandular Tissue Component, GTC)에 따른 Multiview CNN의 성능 변화

선조직성분은 ABUS 영상에서 배경 질질의 밀도와 복잡성을 반영하는 중요한 지표이다. 선조직성분 (Glandular Tissue Component, GTC)가 증가할수록 유선 조직과 섬유

유조직의 비율이 높아지고, 영상 내 구조적 중첩과 음영 변동이 증가하게 된다. 이러한 조건에서는 병변의 경계가 불명확해지거나, 정상 조직과 병변 간의 대비가 감소하여 판독 난이도가 상승한다. 유방 전문의가 아닌 이에서는 최소(minimal)부터 현저한(marked) GTC에 이르기까지 모든 단계에서 AUC가 유의하게 향상되었다. 이는 AI가 전반적인 영상 해석 안정성을 높이는 방향으로 작용했음을 보여준다. 특히 현저한 선조직성분 조건에서 향상 폭이 상대적으로 컸는데, 이는 유선 조직이 풍부한 경우 병변 인식이 더욱 어려워지는 상황에서 AI의 보조 효과가 강조된 결과로 해석된다(그림2).

전문의의 경우 기존 성능이 이미 높은 수준이었으나, 현저한 GTC에서 AUC의 유의한 증가가 관찰되었다. 이는 숙련된 판독자라 하더라도 유선 조직이 매우 치밀한 환경에서는 추가적인 보조 정보가 진단의 확신도를 높일 수 있음을 시사한다. Multiview CNN은 다양한 평면에서의 신호 패턴을 통합하여 조직 간 미세한 구조 차이를 학습하기 때문에, 복잡한 배경 조건에서 병변의 상대적 특징을 보다

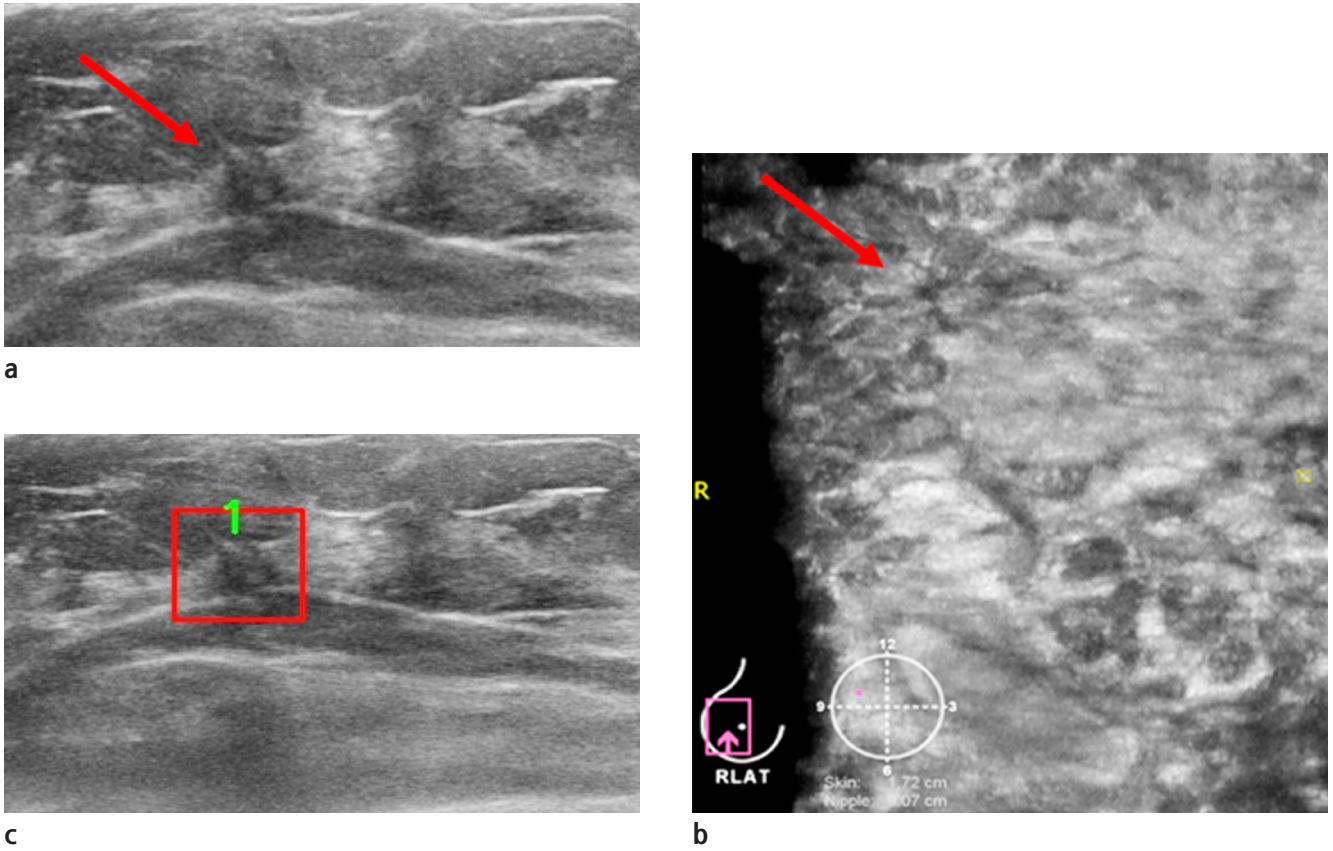


그림 2. 불균질형 배경에코와 현저한 선조직성분을 가진 47세 여성의 ABUS에서 종괴병변으로 침습성 유방암으로 진단된 사례. (a) 횡단면에서 관찰된 종괴 (화살표). (b) 관상면에서 관찰된 동일 병변 (화살표). (c). Multiview CNN이 검출한 병변 (네모 박스). Multiview CNN에 의한 악성 예측도는 0.995였다. Multiview CNN 보조 전과 후의 판독 결과는 다음과 같았다: 유방 영상 전문의가 아닌 이들은 BI-RADS category 4에서 5로 상향 판독하였다. 유방 영상 전문의들은 BI-RADS category 5에서 변화가 없었다.

안정적으로 구분할 수 있는 것으로 보인다.

국내 검진 환경에서 치밀 유방의 비율이 높다는 점을 고려할 때, 현저한 선조직성분 조건에서의 성능 향상은 특히 임상적 의미가 크다. 이는 AI가 단순한 평균적 정확도 향상 도구를 넘어, 치밀 유방이라는 현실적 조건에서 판독의 일관성과 안정성을 높이는 보조 수단이 될 수 있음을 보여준다.

5. 병변 크기에 따른 Multiview CNN의 성능 변화

병변 크기는 유방암 진단에서 중요한 예측 인자일 뿐 아니라, 영상 판독의 난이도에도 직접적인 영향을 미친다. 특히 10mm 이하의 작은 병변은 구조적 단서가 제한적이고, 주변 조직과의 경계가 미세하여 위음성 위험이 상대적으로 높다.

연구 결과, 10mm 이하 병변에서 multiview CNN 보조

후 AUC가 유의하게 향상되었으며, 유방 전문의가 아닌 이와 전문의 모두에서 일관된 개선이 관찰되었다. 특히 작은 병변에서 민감도의 증가가 두드러졌는데, 이는 AI가 미세한 형태학적 차이를 포착하여 위음성을 줄이는 데 기여했을 가능성을 시사한다. 반면 10mm를 초과하는 병변에서는 AI 보조 전후의 성능 차이가 상대적으로 크지 않았다. 이는 비교적 큰 병변은 육안 판독만으로도 충분한 형태학적 단서를 제공하기 때문으로 해석된다.

Multiview CNN은 서로 다른 평면에서의 병변 정보를 통합함으로써 작은 병변의 공간적 연속성과 구조적 왜곡을 보다 명확히 반영할 수 있다. 특히 관상면에서 관찰되는 미세한 조직 당김(retraction)이나 구조적 변형은 작은 악성 병변의 중요한 단서가 될 수 있으며, 이러한 정보의 통합은 조기 암 진단에 기여할 수 있다. 임상적으로 작은 병변의 정확한 분류는 조기 유방암 발견과 직결되며, 불필요한 지연 진단을 줄이는 데 중요한 의미를 가진다. 따라

서 multiview CNN이 작은 병변에서 일관된 성능 향상을 보였다는 점은 조기 검진 환경에서 특히 주목할 만한 결과이다.

결론

ABUS 기반 AI는 최근 딥러닝 기술의 발전과 함께 높은 진단 성능을 보이며 임상적 활용 가능성을 확대하고 있다. 특히 multiview CNN은 서로 다른 평면에서 획득된 정보를 통합함으로써 병변의 형태학적 특징과 공간적 문맥을 동시에 반영할 수 있으며, 단일 뷰 기반 모델에 비해 일관된 성능 향상을 보여주었다. 판독자 경험 수준에 따른 성능 격차를 완화하고, 비종괴 병변, 불균질형 배경 에코, 현저한 선조직성분, 10mm 이하의 작은 병변과 같은 고난도 조건에서 보조 효과가 두드러졌다는 점은 임상적으로 중요한 의미를 가진다.

그러나 평균적인 성능 지표의 향상만으로 실제 임상 통합을 단정하기는 어렵다. 기술적 제약, 데이터 일반화 문제, 적용 방식에 따른 성능 균형, 그리고 검진 환경에서의 실제 workflow 통합과 같은 요소가 함께 고려되어야 한다. 향후 연구에서는 세부 임상 조건에 따른 성능 분석과 다기관 검증을 통해 ABUS 기반 AI의 역할을 보다 구체화할 필요가 있다.

결론적으로, multiview CNN을 포함한 ABUS 기반 AI는 판독을 대체하는 기술이라기보다, 판독자의 의사결정을 보완하고 진단의 일관성을 향상시키는 보조 도구로서 중요한 잠재력을 지닌다. 향후 체계적인 검증과 적용 전략 수립을 통해, 유방 검진 영역에서 보다 안전하고 효과적인 임상 활용이 기대된다.

참고 문헌

1. Siegel RL, Kratzer TB, Giaquinto AN, Sung H, Jemal A. Cancer statistics, 2025. *CA Cancer J Clin*. 2025;75(1):10-45.
2. Scheel JR, Lee JM, Sprague BL, Lee CI, Lehman CD. Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts. *Am J Obstet Gynecol*. 2015;212(1):9-17.
3. Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA*. 2008;299(18):2151-63.
4. Ikeda Y, Fukuoka D, Hara T, et al. Development of a fully automatic scheme for detection of masses in whole breast ultrasound images. *Med Phys*. 2007;34(11):4378-88.
5. Kim SH, Kim HH, Moon WK. Automated Breast Ultrasound Screening for Dense Breasts. *Korean J Radiol*. 2020;21(1):15-24.
6. Han S, Kang HK, Jeong JY, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys Med Biol*. 2017;62(19):7714-28.
7. Komura D, Ishikawa S. Machine learning approaches for pathologic diagnosis. *Virchows Arch*. 2019;475(2):131-8.
8. Ahmad J, Akram S, Jaffar A, et al. Deep learning empowered breast cancer diagnosis: Advancements in detection and classification. *PLoS One*. 2024;19(7):e0304757.
9. Wang Y, Choi EJ, Choi Y, Zhang H, Jin GY, Ko SB. Breast Cancer Classification in Automated Breast Ultrasound Using Multiview Convolutional Neural Network with Transfer Learning. *Ultrasound Med Biol*. 2020;46(5):1119-32.
10. Ph DX, Xu M, Tang G, et al. The lesion detection efficacy of deep learning on automatic breast ultrasound and factors affecting its efficacy: a pilot study. *Br J Radiol*. 2022;95(1130):20210438.
11. Lee SE, Han K, Youk JH, et al. Differing benefits of artificial intelligence-based computer-aided diagnosis for breast US according to workflow and experience level. *Ultrasonography*. 2022;41(4):718-27.
12. Guetari R, Ayari H, Sakly H. Computer-aided diagnosis systems: a comparative study of classical machine learning versus deep learning-based approaches. *Knowl Inf Syst*. 2023:1-41.
13. Ayana G, Dese K, Choe SW. Transfer Learning in Breast Cancer Diagnoses via Ultrasound Imaging. *Cancers (Basel)*. 2021;13(4).
14. Choi JS, Han BK, Ko ES, et al. Effect of a Deep Learning Framework-Based Computer-Aided Diagnosis System on the Diagnostic Performance of Radiologists in Differentiating between Malignant and Benign Masses on Breast Ultrasonography. *Korean J Radiol*. 2019;20(5):749-58.
15. Lee S, Lee HS, Lee E, Kim WH, Kim J, Yoon JH. Improving breast ultrasonography education:

- the impact of AI-based decision support on the performance of non-specialist medical professionals. *Ultrasonography*. 2025;44(2):124-33.
16. Kwon MR, Youn I, Lee MY, Lee HA. Diagnostic Performance of Artificial Intelligence-Based Computer-Aided Detection Software for Automated Breast Ultrasound. *Acad Radiol*. 2024;31(2):480-91.
17. Guo Y, Li N, Song C, Yang J, Quan Y, Zhang H. Artificial intelligence-based automated breast ultrasound radiomics for breast tumor diagnosis and treatment: a narrative review. *Front Oncol*. 2025;15:1578991.
18. Choi EJ, Wang Y, Choi H, et al. Automated breast ultrasound features associated with diagnostic performance of a multiview convolutional neural network according to the level of experience of radiologists. *Ultraschall Med*. 2025.

J Korean Soc Breast Screening 2026;23:12-19

Automated Breast Ultrasound and Artificial Intelligence: Multiview Convolutional Neural Network Performance by Radiologist Experience

Seoyun Choi, Eun Jung Choi

Department of Radiology and Research Institute of Clinical Medicine of Jeonbuk National University-Biomedical Research Institute of Jeonbuk National University Hospital, Jeonbuk National University Medical School, 20 Geonji-ro, Deokjin-gu, Jeonju City, Jeollabuk-Do 54907, South Korea.

Automated breast ultrasound (ABUS) is increasingly used as a supplemental screening modality for women with dense breasts; however, it has several limitations, including prolonged interpretation time and inter-reader variability due to the large volume of three-dimensional data. Recently, artificial intelligence (AI) based on convolutional neural networks (CNNs), particularly multiview CNN strategies that simultaneously utilize transverse and coronal views, has been introduced and shown to improve diagnostic performance. This review summarizes the diagnostic performance of ABUS-based multiview CNN models and analyzes differences in the supportive effect of AI according to reader experience and imaging characteristics, aiming to suggest potential directions for clinical implementation in the Korean screening environment.

Index words: Breast; Automated breast ultrasound; Multiview convolutional neural network; Ultrasound features

Corresponding author: Eun Jung Choi, M.D., Ph.D.