

유방 초음파에서 인공지능의 활용 및 과제

안진경

노원을지대학교병원 영상의학과

유방 초음파는 접근성이 높고 방사선 노출이 없으며 실시간 영상 평가가 가능하여 유방 병변 평가 및 치밀 유방 여성에서의 보조 선별 검사에 널리 사용되고 있다. 그러나 검사자 의존성이 높고 특이도가 상대적으로 낮아 위양성 및 불필요한 생검이 증가하는 한계가 있다. 최근 딥러닝을 기반으로 한 인공지능 기술은 이러한 한계를 보완할 수 있는 도구로 주목받고 있다. 본 리뷰에서는 유방 초음파 인공지능의 발전 현황을 살펴보고, 진단 및 예측 영역에서의 임상적 활용과 실제 임상 도입을 위한 과제를 고찰하고자 한다.

Index words: Breast ultrasound; Artificial intelligence; Deep learning; Computer-aided diagnosis

서론

유방암은 전 세계 여성에서 가장 흔하게 진단되는 악성 종양 중 하나이며, 여전히 주요한 암 관련 사망 원인으로 보고되고 있다. 조기 발견을 통해 전이 이전 단계에서 치료가 이루어질 경우 생존율이 유의하게 향상되므로, 효과적인 영상 진단 전략은 유방암 관리에서 핵심적인 역할을 한다 (1, 2).

현재 유방암 검진의 표준 영상 방법은 유방촬영술이지만, 특히 유방 치밀도가 높은 여성에서는 민감도가 크게 감소하는 한계가 있다. 치밀 유방에서는 유방촬영술의 민감도가 약 48-64%까지 감소하는 것으로 보고되며, 이러한 환자군에서는 보조 영상 검사로서 초음파의 역할이 중요하다 (1).

유방 초음파 (breast ultrasound, US)는 높은 접근성, 이

온화 방사선 노출이 없다는 안전성, 실시간 영상 획득 능력, 그리고 비용 효율성 측면에서 중요한 영상 검사이다. 유방 초음파는 병변 특성 평가, 중재적 시술 유도, 그리고 치밀 유방 여성에서의 보조 선별 검사에 광범위하게 활용되고 있다 (1, 3). 그러나 검사자 의존성이 높고 영상 획득의 변동성이 크며, 유방촬영술 대비 특이도가 상대적으로 낮다는 제한점이 존재한다. 이러한 특성은 위양성 (false-positive) 증가 및 불필요한 생검 (unnecessary biopsy)으로 이어질 수 있다 (1, 4).

최근 딥러닝 (deep learning, DL)을 기반으로 한 인공지능 (artificial intelligence, AI) 기술은 진단의 일관성과 정확도를 향상시킬 수 있는 유망한 도구로 주목받고 있다. 초기에는 컴퓨터 보조 검출 (computer-aided detection, CADe) 및 컴퓨터 보조 진단 (computer-aided diagnosis, CADx)에 주로 적용되었으나, 현재 AI의 활용 범위는 예측 모델링 (prediction modeling)과 위험도 계층화 (risk stratification) 영역으로 확장되고 있다 (1, 4, 5).

본 리뷰에서는 유방 초음파 인공지능의 발전 현황을 살펴보고, 진단 및 예측 영역에서의 임상적 활용과 실제 임상 도입을 위한 과제를 고찰하고자 한다.

통신저자: 안진경
서울시 노원구 한글비석로 68
(01830) 노원을지대학교병원 영상의학과
Tel: 02-970-8290 Fax: 02-970-8346
E-mail: jkan0831@eulji.ac.kr

1. 인공지능 기술 및 상용화 도구

1. 인공지능 주요 용어

영상의학 분야에서 인공지능을 정확히 이해하기 위해서는 기술적 계층 구조와 목적에 따른 용어 구분이 필수적이다. 인공지능 (AI)은 외부 데이터를 정확하게 해석 및 학습하고, 이를 유연하게 적용하여 특정 작업을 수행하는 컴퓨터 시스템의 능력을 총칭하며 인간의 뇌 기능과 유사하게 문제를 해결하는 것을 지향한다. 머신러닝 (Machine Learning, ML)은 인공지능의 하위 분야로, 데이터에서 패턴을 학습하는 알고리즘 프레임워크이다 (6, 7). 과거에는 전문가에 의한 수작업 기반 특징 (handcrafted features) 추출에 의존했으나, 이는 복잡한 패턴 처리에는 한계가 있었다 (1, 7). 딥러닝 (DL)은 다층 신경망 (neural network)을 통해 데이터로부터 추상적인 특징을 스스로 학습하는 기술로서 대규모 데이터를 통해 엔드 투 엔드 (end-to-end) 학습이 가능하며, 인간이 인지하지 못하는 미세한 생물학적 정보까지 추출할 수 있다 (6, 7). 합성곱 신경망 (Convolutional Neural Network, CNN)은 이미지 분석에 특화된 딥러닝 아키텍처로, 픽셀 간의 공간적 관계를 보존하면서 특징을 추출하는 것에 탁월하여 유방 초음파 이미지 해석의 핵심 기술로 자리 잡았다 (6, 7). 라디오믹스 (radiomics)는 의료영상을 고효율의 디지털 정량 정보로 변환하여 종양의 표현형 (phenotype)과 유전적 특성을 분석하는 기술이다 (3, 7). 컴퓨터 보조 검출 (CADe)은 의심

스러운 병변 부위를 시각적으로 표시하여 검출을 돕는 시스템이고, 컴퓨터 보조 진단 (CADx)은 검출된 병변의 악성 가능성을 수치화하거나 분류하여 진단 결정을 보조하는 시스템이다 (4, 8). 생성형 인공지능 (Generative AI, GenAI)은 기존 데이터의 분포를 학습하여 새로운 데이터를 생성하거나 이미지 품질을 향상 (노이즈 제거 등) 시키는 최신 인공지능 분야이다 (7). (표1).

2. 의료영상 인공지능 적용의 발전사

의료영상 AI는 단순한 전산 보조 도구에서 능동적인 의사결정 시스템으로 진화해 왔다. 1956년 존 매카시 (John McCarthy)가 AI 개념을 처음 소개한 이후, 1970년대부터 의료분야 적용이 모색되었다 (6, 7). 초기 시스템은 병변의 모양, 경계 등 육안으로 식별 가능한 특징을 수동으로 코딩하는 방식에 의존했으며, 장비나 프로토콜 변화에 따른 일반화 성능이 낮아 임상적 실효성이 제한적이었다 (1). 2012년 AlexNet이 이미지 인식 대회에서 최고의 성적을 거두면서 딥러닝이 상당한 주목을 받게 되었고, 이는 의료영상 분야에서 자동 특징 추출 기술이 부상하는 계기가 되었다. 영상학과 전문의의 업무 과중과 인력 부족 문제를 해결하기 위해 AI를 활용한 워크플로우 자동화가 본격적으로 연구되었다 (8, 9). 2022년 이후 거대언어모델 (Large Language Model, LLM)과 생성형 AI가 등장하며 판독문 자동 생성 및 멀티모달 (Multi-modal) 데이터 통합이 가능해졌고 (7) 단순 진단을 넘어 생존율 예측, 유전

표 1. 인공지능 용어

Term	Definition	Clinical Application
Artificial Intelligence (AI)	Systems that mimic human intelligence	Automated image interpretation
Machine Learning (ML)	Algorithms that learn patterns from data	Lesion detection and classification
Deep Learning (DL)	Machine learning based on multi-layer neural networks	Automated medical image analysis
Convolutional Neural Network (CNN)	Deep learning architecture designed for image recognition	Lesion detection in US or MMG
Radiomics	Extraction of quantitative features from medical images	Tumor phenotype and prognosis prediction
Computer-Aided Detection (CADe)	AI system that highlights suspicious regions	Calcification or mass detection
Computer-Aided Diagnosis (CADx)	AI system assisting diagnostic decision making	BI-RADS risk assessment
Generative AI (GenAI)	AI models capable of generating new data	Image quality enhancement

체 정보 연계 등 정밀 의료의 핵심 도구로 AI가 기능하기 시작했다 (9).

3. 유방 초음파 인공지능 적용의 발전사

유방 초음파는 검사자 의존도가 높고 위양성률이 높은 특성 때문에 AI의 도입 효과가 매우 유용한 분야 중 하나이다. 2010년대 초반은 전통적 머신러닝 및 특징 추출 단계로서, 초기에는 서포트 벡터 머신 (Support Vector Machine, SVM)이나 랜덤 포레스트 (Random Forest)와 같은 머신러닝 알고리즘을 사용하여 병변의 에코 특징을 분류하려 시도했다 (7). 당시 초음파의 생검 양성예측도 (PPV3)는 9-11%로 낮아 이를 개선하기 위한 AI 기술의 필요성이 대두되었다 (4) 2016년-2020년은 CNN 기반 진단 보조가 보급된 시기이다. 2016년 탄성 초음파 (SWE) 데이터와 딥러닝을 결합하여 종양의 경직도를 분석하는 연구가 좋은 성적을 거두었고, S-Detect (Samsung Medison)와 같은 초음파 장비 내장형 CADx 시스템이 보급되면서 저년차 판독의 숙련도 차이를 보완해 주는 효과를 보였다 (10). 2018년 유방 초음파 영상 (breast

ultrasound images, BUSI)과 같은 공공 데이터셋이 구축되면서 알고리즘 성능 측정을 위한 표준 벤치마크가 마련되었다. 2021년 이후는 다차원적 예측 및 워크플로우 혁신이 이루어지고 있는 시기이다. 자동 유방 초음파 (ABUS)와 AI가 결합하여 방대한 이미지 데이터를 트리아지 (triage)하여 판독 시간을 단축시키는 성과를 보였고 (7), 단순히 암을 찾는 단계를 넘어 분자 아형 (molecular subtype) 예측, 액와 림프절 전이 예측, 수술 전 항암화학요법 반응 예측 등 고도화된 예후 예측으로 연구 영역이 확장되었다 (3, 11-13).

4. 유방 초음파에 사용되는 상용 AI 상품 (표 2)

현재 임상에서 활용 가능한 대표적인 상용 AI 도구들은 크게 진단 지원 (CADx) 또는 검출 지원 (CADE)의 기능으로 구분된다.

S-Detect (Samsung Medison)는 사용자가 병변 중심을 지정하면 알고리즘이 자동으로 병변 경계를 설정하고 BI-RADS 분류 체계에 맞춘 병변의 특징 (모양, 방향, 경계 등)을 분석하여 '양성 가능성' 혹은 '악성 가능성'이라는

표 2. 유방 초음파에 사용되는 상용 AI 상품

제품명 (제조사)	적용영상	워크플로우	출력방식	임상적 활용
S-Detect (Samsung Medison)	HHUS	• 내장형 • 사용자가 병변 중심 지정 • AI가 자동 경계 설정	• 이분법적 결과 (possibly benign / possibly malignant) • BI-RADS descriptor 생성	BI-RADS 기술문 자동화, 검사자 의존성 감소, 불필요한 생검 감소 가능성, 저경력 판독자 보조
CadAI-B (BeamWorks)	HHUS	• 실시간 병변 탐지 및 진단	• 실시간 heatmap (CadAI-Map) • 악성확률 점수 (DadAi-score) • BI-RADS category	실시간 병변 탐지 및 진단 통합, 치밀유방에서 성능 향상 가능성
Koios DS (Koios Medical)	HHUS or ABUS	• 벤더 독립형 • 검사자가 ROI 지정	• 0-1사이의 악성확률 수치 • BI-RADS 2-5등급 (benign / probably benign / suspicious / probably malignant)	정량적 위험도 제공, 판독자 간 일치도 향상
BU-CAD (TaiHao Medical)	HHUS	• 벤더 독립형 • 검사자가 병변 영상 전송 • AI가 병변 윤곽선 생성	• 0-100 사이의 Lesion characteristic score (SLC) • BI-RADS category • BI-RADS descriptor 생성	병변 탐지 및 진단 통합, BI-RADS 기술문 자동 생성, 판독시간 단축
QVCAD (QView Medical)	ABUS	• 3D volume 전체 탐색	• 의심 부위 marking • 지능형 Minimum intensity projection(MIP) navigator 기반 circle overlay	판독 효율 향상, 대용량 영상에서 병변 탐지 보조

HHUS, hand-held ultrasound; ABUS, automated breast ultrasound

이분법적 평가를 제공한다 (4, 8). 메타분석 결과 0.90의 AUC와 82%의 민감도, 83%의 특이도를 나타냈으며, 특히 저년차 영상의학 전문의의 진단 정확도를 전문가 수준으로 보완하는 데 효과적이라고 보고하였다 (4, 7, 8). (그림 1)

CadAI-B (BeamWorks)는 국내혁신기술로서 국내외에서 주목받는 실시간 유방 초음파 진단 보조 시스템이다. 픽셀 단위의 주석 (annotation) 없이도 모델 스스로 악성 부위를 우선 순위화하여 학습하는 '약지도 학습 (weakly supervised learning)' 방식을 채택했다. 실시간 프레임마다 악성 확률과 BI-RADS 등급을 산출하며, 'CadAI-Map'이라는 공간 주의 지도를 통해 병변 내 어느 부위가 악성 특징과 연관되었는지 시각화해 준다. 특히, 치밀 유방 환자군 (Category D)에서 민감도 92.3%, 음성 예측도 96.0%의 높은 성능을 보여, 한국 여성에게 흔한 고밀도 유방 환경에서 특히 유용함을 보여 주었다 (14). (그림 2)

Koios DS (Koios Medical)는 장비 제조사와 무관하게

사용이 가능한 독립형 판독 결정 지원 시스템이다. 정지 영상에서 사용자가 병변을 선택하면 수천 개의 특징점을 분석하여 0에서 1 사이의 악성 확률을 수치로 제공하며, 이를 BI-RADS 최종 평가 등급 (2~5)과 매칭하여 제시한다 (5, 15). 다기관 연구에서 판독의들의 평균 AUC를 0.83에서 0.87로 향상시켰으며, BI-RADS 평가의 판독자 간 편차도 감소하였다 (14, 16). (그림 3)

BU-CAD (TaiHao Medical)는 검출 (CADE)과 진단 (CADx) 기능이 통합된 벤더 독립형 소프트웨어이다. 검사자가 병변 영상을 획득하여 시스템에 전송하면, AI가 자동으로 관심 영역 (ROI)을 생성하고 병변의 윤곽선을 그린다. 0~100 사이에서 병변 특성 점수 (score of lesion characteristics, SLC)를 제공하며, 이에 대응하는 BI-RADS 카테고리과 형태학적 기술문을 함께 산출한다. 판독 시간을 건당 약 40% 단축시키는 효율성 증대 효과를 보고하였다 (4, 8). (그림 4)

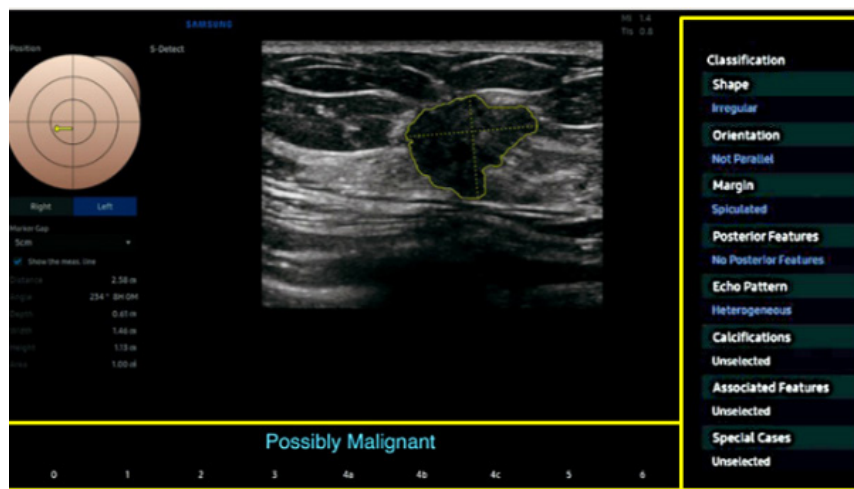


그림 1. S-Detect (Samsung Medison). 검사자가 병변의 중심을 지정하면 S-detect가 자동으로 병변의 윤곽을 그리고 분석한다. 소프트웨어는 최종적으로 병변을 양성 또는 악성으로 분류하고 BI-RADS 범주를 제시한다. 최근 실시간 병변 검출 기능이 추가되었다. Adapted from Brot HF, et al. Ultrasonography 2024.

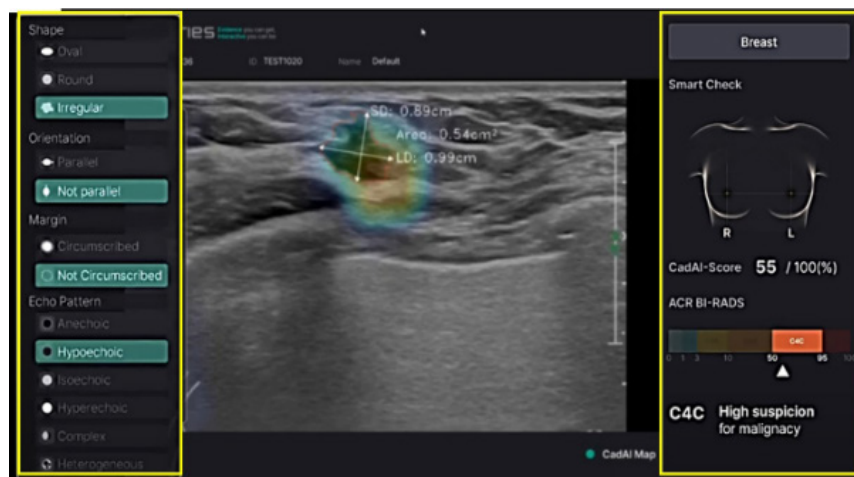


그림 2. CadAI-B (BeamWorks). 실시간 분석으로 잘 알려져 있으며 스캔 중에 병변을 자동으로 감지하고 분석한다. CadAI 점수를 제공하고 병변을 BI-RADS 범주로 분류한다. Adapted from <https://www.youtube.com/@BeamWorks>

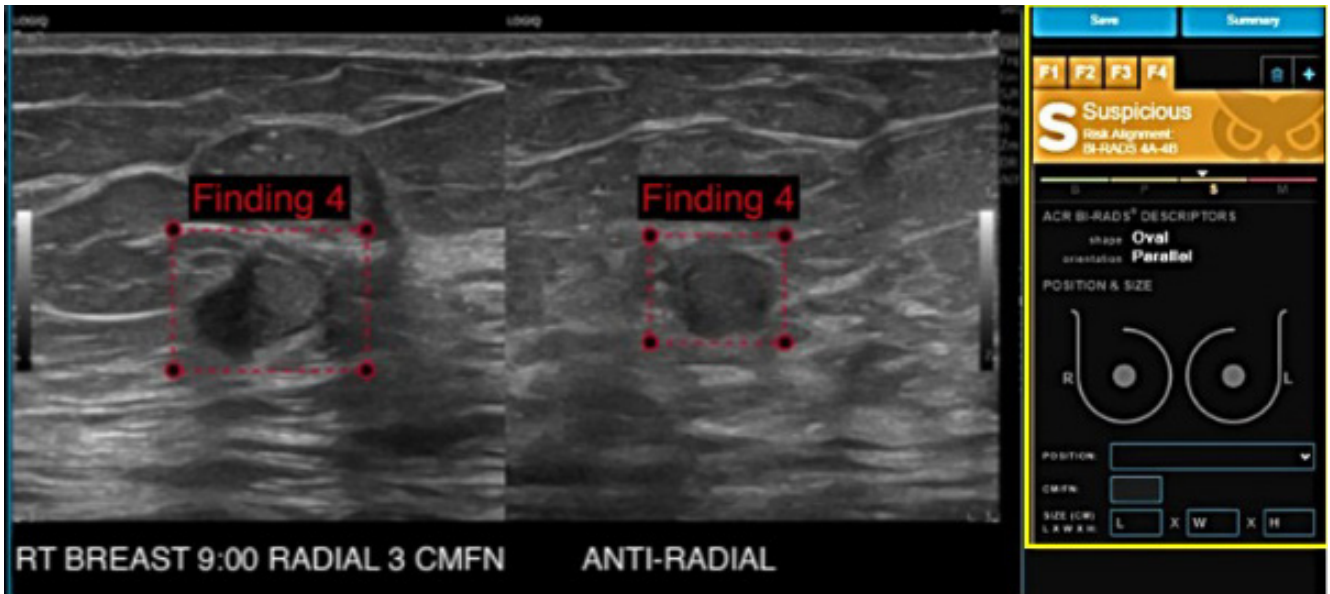


그림 3. Koios DS (Koios Medical). 검사자가 관심 영역 위로 마우스를 드래그하면 소프트웨어가 병변을 분석하고 BI-RADS 범주를 표시한다. Adapted from Brot HF, et al. Ultrasonography 2024.



그림 4. BU-CAD (TaiHao Medical). 검사자가 방사상 (radial)과 역방사상 (antiradial) 영상을 획득한 후 BU-CAD 소프트웨어로 전송한다. 소프트웨어가 자동으로 ROI를 생성하고, 병변 특성 점수 (score of lesion characteristics, SLC) 및 BI-RADS 기술 용어 (descriptor)를 제공한다. Adapted from Brot HF, et al. Ultrasonography 2024.

QVCAD (QView Medical)는 자동 유방 초음파 (ABUS) 전용으로 설계되어 방대한 볼륨 데이터에서 의심스러운 영역을 시각적으로 표시함으로써 판독 시간을 단축시킨다. ABUS가 생성하는 수천 장의 볼륨 데이터에서 5mm 이상의 의심스러운 영역을 자동으로 식별하고 지능형 최소강도 투사 (Minimum Intensity Projection, MIP) 영상에 의심 부위를 원형으로 표시한다. 방대한 ABUS 영상 판독 시 발생하는 시각적 피로도를 줄여주며, 전체 판독 시간을 단축시켜 선별 검사 효율을 높인다고 보고하였다 (7). (그림 5)

II. 유방 초음파에서 인공지능의 임상 적용

1. 진단 영역

1.1 진단 정확도의 향상

최근 연구들은 상용화된 AI 도구들이 숙련된 유방 전문의와 대등하거나 이를 상회하는 성능이 있음을 제시하고 있다. Koios DS는 900개의 병변을 대상으로 한 연구에서 AI 단독 AUROC가 0.88로, 전문의 평균인 0.83보다 높

게 나타났다 (4). S-Detect는 메타분석 결과 0.89의 높은 AUC를 보였으며, 대규모 전향적 연구에서 전문의의 0.74보다 높은 0.83의 AUC를 달성했다 (4, 7, 8). 뉴욕 대학교 (New York University, NYU) 데이터 과학 센터와 NYU 랑곤 헬스 (NYU Langone Health)는 약 28만 건의 방대한 데이터를 학습한 NYU 딥러닝 모델을 개발하였고, 내부 테스트셋에서 AUROC 0.976, 외부 테스트셋에서 AUROC 0.927의 진단 정확도를 보였다. 영상의학과 전문의와의 협업 시 불필요한 생검 요청은 약 27.8% 감소되었다 (1).

1.2 고난도 병변 및 환자군에서의 정밀 진단

인공지능은 인간의 시각적 인지 한계를 보완하여 까다로운 조건에서도 높은 식별력을 유지한다. 유방 밀도가 높을수록 배경 조직의 균일성이 확보되어 민감도 (92.3%)와 음성 예측도 (96.0%)가 더욱 향상되는 특징을 보였고 (14), 크기가 2cm 이하인 소종양이나 진단이 어려운 비종괴성 (non-mass) 병변에서도 MobileNet 등의 경량화 모델을 통해 0.84~0.91의 높은 AUC를 확보할 수 있었다 (7, 15). 인공지능은 특히 불규칙한 모양의 분석에서 인간보다 더 높은 정확도 (74.1% vs. 57.4%, $p = 0.002$)를 보였다 (16).

1.3 멀티모달 통합을 통한 시너지 효과

단일 B-mode 영상의 한계를 극복하기 위해 기능적 정보를 통합하는 멀티모달 AI 기술이 정확도를 향상시키고 있다. 탄성 초음파로 형태학적 정보와 조직 경직도 (stiffness) 정보를 결합했을 때 0.95의 AUC를 얻었고, 경험이 적은 의료진의 진단력을 전문가 수준으로 보완하는데 중요한 역할을 하였다 (3). 도플러 및 조영증강 초음파를 이용하여 혈류 정보나 미세혈관 패턴을 AI로 분석할 경우, 단독 분석보다 AUC가 약 0.08~0.12 가량 상승하여 BI-RADS 4 병변의 감별 진단력을 높여주었다 (7).

1.4 위양성 및 과잉 생검의 감소

정확도가 향상되면 불필요한 침습적 검사를 줄일 수 있다. Shen 등은 인공지능과 전문의의 하이브리드 모델에서 위양성률은 37.3%, 실제 생검 요청은 27.8% 감소하는 효과를 보고하였다 (1). 또 다른 연구에서는 병변의 트리아지를 통해 BI-RADS 4A 병변의 79%를 BI-RADS 3 (probably benign)로 하향 조정하였고, 이중 단 5%만이 악성으로 진단되었다. 반면, AI가 BI-RADS 3에서 4A로 상향 조정한 병변에서는 17%가 악성으로 확인되었다. 이를 통해 암을 놓치지 않으면서도 불필요한 생검을 줄일 수 있다고 보고하였다 (4, 8).

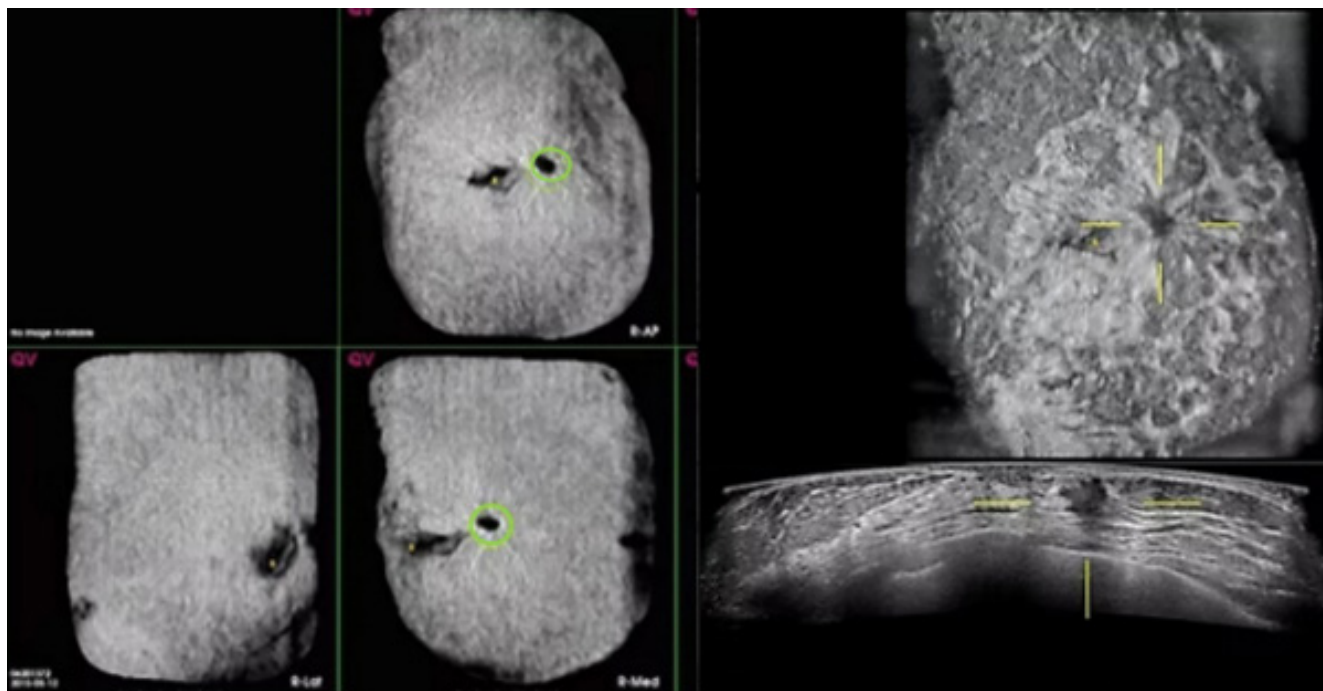


그림 5. QVCAD (QView Medical). QVCAD는 5mm 이상의 의심스러운 부위를 감지하여 녹색 원으로 병변을 표시한다. 검사자가 해당 부위의 관상면 및 횡단면 원본 이미지를 확인할 수 있다. Adapted from <https://www.qviewmedical.com/copy-of-testimonials-1>

1.5 판독 시간의 정량적 단축

상용화된 AI 소프트웨어는 병변의 자동 검출, 크기 측정, 특징 분석 기능을 통해 전문의의 물리적인 판독 시간을 줄여줄 수 있다. 권 등은 최근 발표된 다기관 연구에서 AI 지원 시 건당 중앙 판독 시간이 6.0초에서 3.0초로 50% 단축되었고, 식별력 (discrimination) 지표인 AUROC는 0.921에서 0.953으로 유의미하게 향상 ($p = 0.002$) 되었다고 보고하였다 (15). BU-CAD는 직교 단면 영상 분석 시 판독 시간을 평균 30.15초에서 18.11초로 약 40% 감소시키는 효과를 보여주었다 (4, 8). 이러한 시간 단축 효과는 주니어 판독의 뿐만 아니라 판독량이 많은 상급 종합병원의 숙련된 전문의들에게도 판독 피로도를 낮추고 판독 과정의 질적 안정성과 속도를 개선하는데 실질적인 혜택을 제공할 수 있다 (4, 17).

1.6 신속한 트리야지 및 정상 사례 분류

모든 사례를 동일한 비중으로 검토하는 대신 AI가 저위험군을 신속하게 분류하면, 전문의가 고위험 병변에 더 많은 시간을 할애할 수 있다. NYU의 대규모 데이터 기반 연구에서 AI 시스템은 99.86%라는 매우 높은 음성 예측도 (NPV)를 유지하면서 정상 및 양성 사례의 77.7%를 판독 목록에서 안전하게 제외할 수 있는 잠재력을 보여 주었다. 불필요한 생검 요청의 감소는 영상의학과 및 병리과의 후속 업무 부하의 경감에도 기여할 수 있다 (1).

1.7 기술 문장의 표준화 및 자동 보고

AI는 BI-RADS 분류 체계에 따른 병변의 형태학적 기술 (descriptors)을 객관적으로 제시하여 보고서 작성의 효율성을 높일 수 있다. 결과지 자동 생성의 예로, S-Detect나 Koios DS와 같은 시스템은 병변의 모양, 방향, 경계 등을 자동으로 분석하여 결과지에 즉각 반영할 수 있는 표준화된 기술 문장을 제공한다 (4, 8). 이는 판독자 간 편차 감소의 효과로 숙련도가 낮은 판독의들의 판독 일관성을 높이고 추가적인 컨설팅이나 재판독에 소요되는 시간을 줄이는데 기여할 수 있다 (6, 8). 2025년 이후 부상한 거대언어모델 기반의 생성형 AI는 멀티모달 데이터를 통합하여 판독문의 초안을 자동 생성하는 수준에 도달했으며, 이는 전문의의 행정 업무를 경감시킬 수 있다 (18, 19).

2. 예측 영역

2.1 분자 아형 예측

인공지능 적용의 임상적 이점은 종양 전체를 분석할 수

있다는 점이다. 침상 생검은 종양의 일부분만을 채취하기 때문에 종양 내부의 유전적 이질성으로 인해 실제 아형을 완벽히 대표하지 못할 위험이 있다 (8, 12). 딥러닝 모델 (예: MDL-IIA)은 유방 촬영술과 초음파 영상 전체를 분석하여 종양 내부의 미세한 질감과 패턴을 파악함으로써, 생검이 놓칠 수 있는 하위 클론 (subclones)의 특성을 반영한 포괄적인 아형 예측을 가능하게 한다 (12).

또한, HER2 2+와 같이 임상 현장에서 해석이 까다로운 경계성 결과에도 딥러닝의 도움을 받을 수 있다. 최신 모델들은 면역조직화학 (Immunohistochemistry, IHC) 결과의 불확실성을 극복하기 위해 은염적정법 (Silver In Situ Hybridization, SISH)이나 형광인적정법 (fluorescence in situ hybridization, FISH)과 같은 유전자 수준의 'Ground Truth'를 직접 학습한다. 연구에 따르면 딥러닝 모델은 HER2 2+ 사례에서 약 86.6%의 높은 정확도를 보였으며, 이는 기존의 숙련된 전문의들이 영상만으로 예측하는 수준을 유의미하게 상회한다 (12).

딥러닝은 단일 영상 분석을 넘어 다양한 영상 정보를 통합하는데 매우 유용하다. 멀티모달 AI는 병변의 모양, 경계 등 형태학적 특징에 집중하는 단일 B-mode 영상에 기능적 정보를 더한다 (8). 예를 들어, 도플러는 종양의 혈관 구조와 혈류 정보를 제공하고, 탄성 초음파는 조직의 경직도를 정량화한다 (11). B-mode의 형태학적 정보, 도플러의 혈류 정보, 탄성 초음파의 조직 경직도 정보를 동시에 처리하는 ACNN (Assembled CNN) 모델은 단일 모달리티 모델보다 훨씬 우수한 AUC (0.89-0.96)를 기록하고 있다 (11). 일부 멀티모달 딥러닝 모델은 아형 분류 성능에서 수술 전 생검 (CNB) 결과보다 더 높은 정확도 (AUC 0.89-0.99 vs. 0.67-0.82)를 보여 주었다 (11).

2.2 선행 항암 화학요법 반응 예측

선행 항암 화학요법은 국소 진행성 유방암의 병기를 낮추고 유방 보존술의 가능성을 높이기 위해 시행된다 (3, 13). 하지만 환자마다 반응이 매우 상이하며, 효과가 없는 환자의 경우 불필요한 독성 부작용을 겪으며 적절한 치료 시기를 놓칠 위험이 있다 (13). 기존 방식은 종양의 크기 변화 (Response Evaluation Criteria in Solid Tumors, RECIST)를 기준으로 측정하지만, 치료에 반응하여 종양이 섬유성 조직으로 대체될 경우 크기 변화가 거의 없어 반응 여부를 정확히 판단하기 어려운 경우가 많다 (4, 8). 또한, 검사자의 주관에 따라 결과가 달라지는 문제도 존재한다 (3). AI 및 딥러닝 라디오믹스 (DLR)의 적용으로 병리적 완전 반응 (pCR)을 예측할 수 있는데, Jiang 등은 치료

전후의 초음파 데이터를 통합한 딥러닝 라디오믹스 노모그램 (DLRN)을 개발하여 pCR을 예측하는 데 AUROC 0.94라는 높은 성과를 거두었다. 이는 숙련된 전문의의 육안 판독보다 우수한 수치이다 (3, 8). 조기 반응 예측의 연구로 Gu 등은 항암 2차 및 4차 주기 후의 초음파 영상을 분석하는 단계적 예측 모델을 제안했다. 특히 2차 주기 직후에 비반응 환자의 약 90%를 성공적으로 식별해냄으로써, 치료 초기 단계에서 효과 없는 약물을 중단하거나 전략을 수정할 수 있는 임상적 근거를 마련하였다 (13). 임상적 의의는 AI가 단순한 크기 변화를 넘어 종양 내부 질감과 미세환경의 생물학적 변화를 픽셀 단위로 분석하여 초기에 반응 정보를 제공함으로써 환자 맞춤형 치료 계획 수립을 보조할 수 있다는 점이다 (3, 13).

2.3액와 림프절 전이 예측

액와 림프절 상태는 유방암의 병기 결정과 수술 범위 선택 (감시 림프절 생검 vs. 액와 림프절 광청술)에 결정적인 역할을 한다 (2, 20). 기존 방식의 경우, 액와부 초음파를 통한 직접적인 관찰에서 전이 림프절이 뚜렷하지 않은 경우 (clinically node-negative) 민감도가 낮으며, 특히 육안으로 보이지 않는 미세 전이 (micrometastasis)를 놓칠 확률이 15~20%에 달한다. 최근의 AI 기술은 원발 부위 기반 예측으로, 전이 여부를 판단하기 위해 림프절 자체를 보는 것이 아니라 원발 종양 (primary tumor)의 초음파 특징을 분석한다 (4, 20). 종양의 경계부 패턴과 내부 저에코 영역의 특징이 림프절 전이와 밀접한 상관관계가 있음을 딥러닝 모델이 스스로 학습한 결과이다 (2, 20). Zheng 등의 연구에 따르면, 일반 초음파 영상에 탄성 초음파와 임상 데이터를 결합한 모델은 전이 유무뿐만 아니라 전이된 림프절의 개수 (heavy vs. low burden)까지 구분해냈고 0.905의 AUC를 보였다 (2). 이는 ACOSOG Z0011 임상 시험 결과에 따라 불필요한 액와부 광청술을 피하고 수술 후 림프부종과 같은 부작용을 최소화하려는 현대 유방 외과의 흐름과도 일치한다. 저자는 림프절 전이 상태를 예측하는 AI 모델이 비침습적 영상 바이오마커 (biomarker)로서 가능성이 있음을 제시하였다 (2).

2.4 초기 치료 전략 수립 및 워크플로우 최적화

신속한 트리아지는 생검 결과가 나오기 전이라도 영상 분석만으로 고위험 아형 (예: TNBC) 가능성을 초기에 시사하여, 선행 항암 화학요법이나 수술 계획을 보다 빠르게 논의할 수 있게 한다 (4, 8). 아형 예측뿐만 아니라 Ki-67 증식 지수나 액와 림프절 전이 상태와 연계된 통합적인 예

후 정보를 제공함으로써 진정한 의미의 맞춤형 정밀 의료 (personalized medicine) 구현을 돕는다 (4, 9, 12).

III. 인공지능 도입을 위한 과제

1. 기술적 및 실무적 제약

1.1 데이터의 희소성 및 편향성

유방 초음파 AI 모델 학습을 위한 고품질의 정제된 데이터셋은 유방촬영술에 비해 현저히 부족한 실정이다 (4, 10). 현재 대부분의 연구는 특정 단일 기관의 데이터에 의존하는 경향이 있고, 다양한 인종적·지리적 특성을 반영하지 못해 알고리즘의 보편성을 확보하는 데 한계가 있다 (4, 9). 또한, 임상 지표 외에 유전자 데이터와 같은 개인화된 정보를 통합 수집하는 것 역시 기술적으로 매우 까다로운 과제이다 (18).

1.2 도메인 시프트 (domain shift)와 일반화 (generalizability) 성능

유방 초음파는 전 세계적으로 10개 이상의 주요 제조사가 각기 다른 트랜스듀서 기술과 설정을 사용하고 있다 (10). 도메인 시프트는 인공지능 모델이 학습 과정에서 경험한 데이터 분포 (source domain)와 실제 임상에서 마주하는 데이터 분포 (target domain)가 일치하지 않는 경우 모델의 예측 성능이 저하되는 현상을 의미한다. 따라서, 특정 장비로 학습된 AI 모델이 다른 제조사의 장비나 다른 의료기관의 환경에서 성능이 급격히 저하되는 도메인 시프트를 겪을 수 있으며, 이는 임상 도입의 큰 기술적 걸림돌 중 하나이다 (7).

1.3 검사자 의존성 및 표준화 문제

수동 초음파 (Hand-held ultrasound, HHUS)는 검사자의 숙련도와 압박 정도에 따라 영상의 질이 크게 달라지는 검사자 의존성을 가진다 (4, 8). 이러한 비표준화된 영상 획득 방식은 AI의 진단 일관성을 해칠 수 있으므로, 하드웨어와 무관하게 일정한 성능을 낼 수 있는 보편적인 알고리즘 개발과 영상 획득 프로토콜의 표준화가 필요하다 (7).

1.4 워크플로우 통합 및 블랙박스 문제

AI 솔루션이 판독 효율을 높이려면 기존의 PACS (영상 전송 시스템) 환경에 매끄럽게 통합되어야 하지만 아직 많은 제품이 별도의 워크스테이션이나 수동 ROI 설정을 필요로 한다 (7, 9, 10). 또한, 딥러닝 모델의 복잡한 의

사결정 과정을 인간이 직관적으로 이해할 수 없는 블랙박스 문제는 의료진의 신뢰를 저해하는 요소이기도 하다 (1, 9). 의료진과 환자가 AI를 신뢰하기 위해서는 설명 가능성 (explainability)도 확보되어야 한다 (7, 9). Grad-CAM (Gradient-weighted Class Activation Mapping)과 같은 시각화 도구를 통해 AI가 종양의 어느 부위를 근거로 악성을 의심했는지 제시함으로써 임상적 신뢰를 쌓는 과정이 필요하다 (7, 9, 10). 또한, 실제 임상 워크플로우 내에서 AI 성능 저하 여부를 지속적으로 평가하기 위한 사후 모니터링 체계와 거버넌스 구축이 필요하다 (4, 7).

2. 윤리적 및 개인정보 보호

AI 모델의 고도화를 위해서는 대규모의 고품질 의료 데이터가 필수적이지만, 이 과정에서는 윤리적 및 법적 문제가 발생할 수 있다.

2.1 데이터 개인정보 보호 및 보안

민감한 의료 데이터의 수집 및 공유 과정에서 개인정보 침해 우려가 지속적으로 제기되고 있다 (6, 9). 특히 다기관 협력을 통한 데이터 공유는 알고리즘의 일반화 성능을 높이는 데 필수적이지만, 개인정보 보호 규제로 인해 데이터 중앙 집중화가 어려운 실정이다. 이를 해결하기 위해 최근에는 원본 데이터를 이동시키지 않고 분산된 환경에서 학습을 진행하는 '연합 학습 (federated learning)' 프레임워크 등이 대안으로 모색되고 있다 (10).

2.2 AI 모델의 편향성

현재 다수의 AI 모델은 특정 인종이나 특정 연령층의 데이터에 치우쳐 학습되는 경향이 있다. 이러한 데이터 편향성은 소수 집단이나 소수 인종 환자에게 대한 진단 정확도를 떨어뜨려 의료 서비스의 형평성을 저해하고 건강 격차를 심화시킬 우려가 있다 (4, 9). 따라서 모델 개발 단계부터 지리적, 인종적으로 다양한 데이터셋을 확보하려는 노력이 필요하다 (4).

2.3 책임 소재

인공지능이 진단 보조를 넘어 독자적인 판단을 내릴 가능성이 제기되면서, 오진 발생 시 책임 소재에 대한 논의가 활발하다 (4, 9). 현재 모든 의료 영상 판독의 법적 책임은 최종 판독을 수행한 영상의학 전문의에게 있다 (4). AI 모델이 블랙박스처럼 작동하여 판단 근거를 명확히 제시하지 못할 경우, 의사가 AI의 제안을 맹목적으로 수용하는 '자동

화 편향 (automation bias)'에 빠질 위험이 있으며, 이는 환자 안전과 직결된 윤리적 문제이다 (1, 4).

3. 규제 및 거버넌스 과제

기술적 완결성 못지않게 이를 사회적으로 수용하고 제도화하기 위한 거버넌스 구축이 필요하다.

3.1 법적 및 윤리적 준수

AI가 의료기기로서 임상에 배치되기 전에 규제 기관 (FDA, CE 등)의 승인을 받기 위한 엄격한 임상 검증 절차가 필수적이다 (6, 9). 최근의 규제 가이드라인은 AI를 '의료 소프트웨어'로 정의하며, 데이터의 투명성, 알고리즘의 안정성, 그리고 지속적인 성능 모니터링을 요구하고 있다 (7).

3.2 가이드라인 수립

유방 초음파의 강한 검사자 의존성을 극복하기 위해 영상 획득 및 라벨링에 대한 범용적인 표준 프로토콜이 수립되어야 한다. 더 나아가 AI의 정량적 리스크 점수를 기존의 BI-RADS 보고 체계에 어떻게 유기적으로 통합할 것인지에 대한 구체적인 임상 가이드라인이 마련되어야 한다. 이를 위해 공학자, 임상, 규제 당국 간의 다학제적 협력이 거버넌스 차원에서 지속되어야 할 것이다 (6, 7).

4. 경제적 비용 및 도입 장벽

4.1 막대한 비용 및 의료 격차 심화

AI 시스템 도입은 단순히 소프트웨어 구매를 넘어 고성능 GPU 하드웨어 구축과 IT 인프라 확장을 위한 상당한 자본 투자를 요구한다. 또한 지속적인 알고리즘 업데이트와 성능 모니터링을 위한 운영 비용도 고려해야 한다 (9). 고가의 AI 솔루션은 자본이 풍부한 대형 병원 위주로 도입될 가능성이 있으며, 결과적으로 저소득 지역이나 자원이 부족한 환경 (low-resource setting)과의 의료 질 격차를 더욱 심화시킬 우려가 있다 (8, 9). 특히 소수의 데이터로 개발된 AI 모델은 소수 민족이나 특정 연령층을 소외시키는 알고리즘 편향성을 유발하여 보건의료의 형평성을 해칠 수 있다 (9).

4.2 건강보험 수가 및 보상 체계의 부재

현재 대다수의 국가에서 AI를 이용한 영상 판독에 대해 명확한 건강보험 수가를 책정하지 않고 있다 (4, 7). 미국에서는 비탄성 초음파 정량 분석을 위한 CPT 코드

(0689T, 0690T)가 활용되기 시작했으나, 지불 주체마다 기준이 상이하어 지속 가능한 수익 모델을 구축하는 데 한계가 있다 (4). 경제적 보상이 뒷받침되지 않는 상황에서 AI 도입은 병원의 일방적인 비용 지출로 간주될 위험이 있다 (9).

4.3 교육 자원 및 신뢰 구축의 비용

AI 도구를 안전하고 효과적으로 사용하기 위해서는 판독의들이 AI의 한계를 이해하고 결과에 대해 주체적인 판단을 내릴 수 있는 인공지능 리터러시 (literacy) 교육이 필요하다 (4). 하지만 이러한 전문 교육 시스템 구축과 임상적 신뢰를 확보하기 위한 전향적 다기관 임상시험에는 막대한 시간과 인적 자원이 소모되며, 이는 또 다른 형태의 도입 장벽으로 작용할 수 있다 (7, 9).

결론

유방 초음파에서 인공지능은 특이도 향상, 위양성 감소, 불필요한 생검 감소 측면에서 의미 있는 임상적 잠재력을 보여주고 있다. 분자 아형, 선행 항암 화학요법, 림프절 전이 등 예측 영역으로의 확장은 개인 맞춤 치료 (personalized management)에 기여할 가능성을 제시한다. 기술적 정밀도만큼 중요한 것은 실제 임상 환경에서의 지속 가능성이다. 인공지능이 단순한 기술적 혁신을 넘어 표준 의료 행위로 안착하기 위해서는 전향적 검증 (prospective validation), 데이터 다양성 확보, 워크플로우 통합, 설명 가능성 강화가 필요하다. 또한 합리적인 수가 체계 마련, 의료 격차 해소를 위한 거버넌스 구축, 그리고 윤리 및 규제 프레임워크의 정립이 병행되어야 한다.

References

1. Shen Y, Shamout FE, Oliver JR, et al. Artificial intelligence system reduces false-positive findings in the interpretation of breast ultrasound exams. *Nat Commun*. 2021;12(1):5645.
2. Zheng X, Yao Z, Huang Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun*. 2020;11(1):1236.
3. Jiang M, Li CL, Luo XM, et al. Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant

- chemotherapy in locally advanced breast cancer. *Eur J Cancer*. 2021;147:95 – 105.
4. Bahl M, Chang JM, Mullen LA, Berg WA. Artificial Intelligence for Breast Ultrasound: AJR Expert Panel Narrative Review. *AJR Am J Roentgenol*. 2024;223(6):e2330645.
5. Berg WA, López Aldrete AL, Jairaj A, et al. Toward AI-supported US Triage of Women with Palpable Breast Lumps in a Low-Resource Setting. *Radiology*. 2023;307(4):e223351.
6. Al-Karawi D, Al-Zaidi S, Helael KA, et al. A Review of Artificial Intelligence in Breast Imaging. *Tomography*. 2024;10(5):705 – 26.
7. Ma T, Wang Z, Dong J, Cheng Y, Zhao H, Cui X. Ultrasound-based artificial intelligence for breast lesion classification. *Front Oncol*. 2026;16.
8. Fruchtmann Brot H, Mango VL Artificial intelligence in breast ultrasound: application in clinical practice. *Ultrasonography*. 2024;43(1):3 – 14.
9. Ahn JS, Shin S, Yang SA, et al. Artificial Intelligence in Breast Cancer Diagnosis and Personalized Medicine. *J Breast Cancer*. 2023;26(5):405 – 35.
10. Kim J, Kim HJ, Kim C, Kim WH. Artificial intelligence in breast ultrasonography. *Ultrasonography*. 2021;40(2):183 – 90.
11. Zhou BY, Wang LF, Yin HH, et al. Decoding the molecular subtypes of breast cancer seen on multimodal ultrasound images using an assembled convolutional neural network model: A prospective and multicentre study. *EBioMedicine*. 2021;74:103684.
12. Zhang T, Tan T, Han L, et al. Predicting breast cancer types on and beyond molecular level in a multimodal fashion. *NPJ Breast Cancer*. 2023;9(1):16.
13. Gu J, Tong T, He C, et al. Deep learning radiomics of ultrasonography can predict response to neoadjuvant chemotherapy in breast cancer at an early stage of treatment: a prospective study. *Eur Radiol*. 2022;32(3):2099 – 109.
14. Lee J, Kim WH, Kim J, et al. Effect of a Real-Time Artificial Intelligence-Assisted Ultrasound System on BI-RADS C4 Breast Lesions Based on Breast Density. *Cancers (Basel)*. 2026;18(3).
15. Kwon H, Kim SM, Oh SH, et al. Artificial Intelligence – Assisted Breast Ultrasound: Modest AUROC Improvement and Shorter Interpretation

- Time without Significant Change in Diagnostic Accuracy. *Frontiers in Radiology*. 6:1747783.
16. Heller SL, Wegener M, Babb JS, Gao Y. Can an Artificial Intelligence Decision Aid Decrease False-Positive Breast Biopsies? *Ultrasound Q*. 2020;37(1):10–5.
17. Tsuyuzaki S, Fujioka T, Yamaga E, et al. Evaluation of AI diagnostic systems for breast ultrasound: comparative analysis with radiologists and the effect of AI assistance. *Jpn J Radiol*. 2025;43(10):1643–51.
18. Liu J, Pian L, Chen J, et al. Artificial intelligence in breast ultrasound: a systematic review of research advances. *Front Oncol*. 2025;15:1619364.
19. Azhar K, Lee BD, Byon SS, Lee S, Cho KR, Song SE. Semiautomated breast ultrasound report generation using multimodal large language models and deep learning. *Front Med (Lausanne)*. 2026;13:1679203.
20. Zhou LQ, Wu XL, Huang SY, et al. Lymph Node Metastasis Prediction from Primary Breast Cancer US Images Using Deep Learning. *Radiology*. 2020;294(1):19–28.

J Korean Soc Breast Screening 2026;23:1-11

Applications and Challenges of Artificial Intelligence in Breast Ultrasound

Jin Kyung An

Department of Radiology, Nowon Eulji University Hospital, Eulji University School of Medicine

Breast ultrasound is widely used for the evaluation of breast lesions and as an adjunctive screening modality in women with dense breasts, owing to its high accessibility, lack of ionizing radiation, and capability for real-time image assessment. However, it is inherently operator-dependent and has relatively low specificity, which can lead to increased false-positive findings and unnecessary biopsies. Recently, deep learning-based artificial intelligence (AI) has emerged as a promising tool to overcome these limitations. In this review, we provide an overview of recent advances in artificial intelligence for breast ultrasound and explore its clinical applications in diagnostic and predictive settings, along with the challenges associated with its integration into real-world clinical practice.

Index words: Breast ultrasound; Artificial intelligence; Deep learning; Computer-aided diagnosis

Corresponding author: Jin Kyung An, M.D., Ph.D.