

유방촬영술 기반 유방암 위험 예측: 최신 딥러닝 모델과 설명 가능한 인공지능 방법론

정수민¹ · 배민선²

¹고려대학교 안산병원 의생명연구센터, ²고려대학교 안산병원 영상의학과

서 론

유방암은 전 세계 여성에게 가장 흔히 발생하는 암으로, 조기 발견을 위해서는 개인별 위험도를 정확하게 평가하는 것이 중요하다. 널리 사용되는 전통적인 방법인 Gail 모델 및 Tyrer-Cuzick 모델은 임상 정보(나이, 가족력 등)를 바탕으로 5년 또는 10년 이내 유방암 발생 위험을 추정하지만, 예측 정확도(area under the curve; 이하 AUC)가 0.6 내외로 제한적이고 개인별 특성을 충분히 반영하지 못한다(1-3). 특히 위험 인자가 없더라도 많은 유방암 환자가 발생하는 현실은 새로운 접근법의 필요성을 보여준다(4). 최근 딥러닝(deep learning)을 이용해 유방촬영술에서 미묘한 이상 소견을 감지하여 암 발생 위험을 직접 예측하려는 연구들이 주목받고 있다. 이번 종설 논문에서는 영상 기반 위험도 예측 기술의 발전을 살펴보고, 특히 Mirai 모델과 이를 설명 가능한 인공지능(explainable AI; 이하 XAI)으로 확장한 최신 연구를 중점적으로 소개한다. 또한 기존 임상 기반 위험 모델과 영상 기반 AI 모델의 구조 및 성능을 비교하고, 영상의 해석 가능성을 높이기 위한 최신 XAI 방법론과 그 임상적 활용 가능성을 포괄적으로 검토한다.

1. 유방암 발생 위험도 예측 모델의 발전

1.1 전통적 위험 요인 기반 모델

기존의 유방암 위험도 예측 도구는 역학적 연구로 밝혀진 위험 요인을 조합하여 발병 확률을 계산한다. 가장 대표적인 Gail 모델은 나이, 첫 월경 나이, 출산 이력, 가족력 등을 기반으로 유방암 위험을 예측하며, 간편하고 널리 활용되지만 예측 성능이 낮은 편이다(1, 3). 가족력과 유전적 요소를 보완한 Tyrer-Cuzick 모델 역시 성능이 제한적이고, 영상 정보는 유방 밀도(breast density)만 반영된다(3). 유방촬영술에서 미세석회화나 비대칭성과 같은 잠재적 영상 소견들은 위험도 평가에 반영되지 않고 있다(5).

1.2 영상 기반 접근과 딥러닝의 도입

영상 정보를 이용한 위험도 평가 연구는 초기에 유방촬영술에서 유방실질조직 패턴이나 질감(texture) 분석으로 시작되었지만, 충분한 예측력을 달성하지 못했다. 그러나 2010년대 이후 대규모 디지털 유방촬영 데이터와 딥러닝 기술의 발전이 만나면서 본격적으로 영상 기반 위험 예측이 가능해졌다. 2019년 Yala 등은 약 8만 9천장의 유방촬영 영 검진 영상을 이용해 딥러닝 분석으로 5년 내 유방암 발생 위험을 예측하는 모델을 개발하여 AUC 0.68을 보고하였다. 이는 기존 Gail 모델이나 Tyrer-Cuzick 모델(AUC 0.59-0.62)보다 유의하게 높았으며, 임상 정보를 추가한 혼합 모델에서는 AUC가 0.70까지 개선되었다(6). 같은 시기 Portnoi 등은 유방 자기공명영상(MRI) 단독으로 5년 위험도를 예측하여 Tyrer-Cuzick 모델보다 우수한 성능(AUC 0.64 vs 0.49)을 입증하였다(7). 이와 같은 초기 성

이 종설은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 이루어진 것입니다(RS-2024-00361049).

통신저자: 배민선

(15355) 경기도 안산시 단원구 적금로 123,

고려대학교 안산병원 영상의학과

Tel. +82-31-412-5228 Fax. +82-31-412-5224

E-mail: minsunb@gmail.com

과 이후 많은 연구 기관에서 유방영상과 딥러닝을 결합한 위험 모델들을 개발하여 예측 성능과 일반화 능력을 지속적으로 향상시키고 있다.

2. Mirai: 유방촬영술 기반 유방암 위험 예측 모델

2.1 모델 구조 및 입력

Mirai는 미국의 매사추세츠 공대(MIT)와 하버드 의대 부속병원(MGH) 연구팀이 개발한 최신 딥러닝 모델로, 유방촬영술 영상만으로 5년 내 유방암 발생 위험도를 예측한다(8, 9). Mirai 모델은 양측 유방에서 각각 2장의 표준촬영 이미지(craniocaudal view [CC], mediolateral oblique view [MLO])를 입력받아 사전 학습된 합성곱 신경망(convolutional neural network; 이하 CNN)으로 영상의 고차원적인 특징을 추출한다. 이후 트랜스포머(transformer)를 이용해 각 영상 간의 상관 관계와 시간 정보를 결합하고, 최종적으로 1~5년의 연차별 위험도를 산출한다(그림 1). 유방암 가족력과 같은 임상 정보가 누락된 경우에도 이 모델은 평균값을 활용하여 예측의 안정성을 유지할 수 있다(9).

2.2 예측 성능 및 기존 모델과의 비교

Mirai 모델은 미국 MGH 데이터를 사용하여 개발되었는데, 스웨덴 카롤린스카(Karolinska) 대학병원(C-index=0.81)과 대만 CGMH(Chang Gung Memorial Hospital) 병원(C-index=0.79) 등 외부 데이터에서도 높은 예측 성능을 보이며 기존의 Tyrer-Cuzick 모델(C-index=0.64)을 능가했다(8). 특히 다양한 인종 및 연령 집단에서 일관된 성능을 보이며, 상위 10% 고위험군의 예측 정확도가 Tyrer-Cuzick 모델 대비 약 1.8배 높았다(9). 최근 영국 국민보건서비스(NHS) 검진 데이터에서도 Mirai는 AUC 0.67-0.68으로 유방밀도만을 이용한 모델(AUC 0.56) 대비 우수한 성능을 보였으나, 영상 장비 제조사 간 성능 차이가 보고되었다(10, 11).

2.3 임상 적용을 위한 과제

Mirai 모델은 뛰어난 예측 성능에도 불구하고 블랙박스 구조로 인해 임상 현장에서 예측 근거를 이해하기 어려운 한계가 있다(12). 이로 인해 의료진의 신뢰 확보와 환자 상담의 활용성이 제한된다. 이를 극복하기 위해 Mirai의 예

측 과정을 보다 명확하게 해석할 수 있는 방법론이 요구되며, 최근 연구에서 설명 가능한 AI 모델들이 등장하였다.

3. Mirai의 해석 가능성과 파생 모델들

3.1 AsymMirai: 유방의 비대칭성을 이용한 설명 가능 모델

2024년 Donnelly 등은 Mirai의 예측 논리를 탐구하고 해석성을 높인 대안 모델로 AsymMirai를 개발하였다(12). Mirai의 복잡한 CNN 및 트랜스포머(transformer) 구조로 인해 일반적인 사후 설명 방법(Grad-CAM 등)이 적합하지 않음을 지적하고, 양측 유방 간 국소적 비대칭성(bilateral dissimilarity)에 주목했다. 즉, Mirai가 위험 예측 시 좌우 유방조직 간의 미세한 차이를 중요한 단서로 활용할 수 있다는 가설을 세우고 이를 모델 구조에 반영하였다. AsymMirai는 Mirai와 동일한 CNN 기반 특징 추출을 사용하되, 복잡한 트랜스포머 모듈 대신 양측 유방 영상 간의 국소적인 차이를 직접 계산하는 구조를 채택하였다(12). 구체적으로는 좌우의 각 뷰(CC, MLO)에서 CNN 특징맵을 위치별로 비교하여, 차이가 가장 큰 영역을 예측 윈도우(prediction window)로 정의한다. 이후 각 뷰별 최대 차이값을 결합한 위험 점수(risk score)를 산출하여 1~5년 위험 확률을 최종적으로 예측한다. 또한 이 구조는 원본 영상에 결과를 중첩(overlay)하여, 좌우 차이가 발생한 부위를 시각적으로 명확히 제시할 수 있다(그림 1).

대규모 EMBED(EMory BrEAST Imaging Dataset) 데이터셋(21만여 건)을 이용하여 AsymMirai의 유방암 발생 예측 성능을 평가한 결과, 1년 AUC 0.79, 5년 AUC 0.66으로 Mirai(각각 0.84, 0.71)의 성능과 근접하였다. 두 모델 간 예측값 상관관계수도 0.68 이상으로 높아, Mirai의 예측력 상당 부분을 비대칭성만으로 재현할 수 있었다(12). AsymMirai의 AUC가 Mirai보다 약간 낮아 통계적으로 유의한 차이를 보였으나, 임상적으로 매우 의미있는 결과가 관찰되었다. 예를 들어, 연속 촬영에서 매년 동일한 부위를 비대칭성으로 강조한 환자군은 실제 유방암 발생률이 매우 높았으며 3년 위험 예측 AUC는 0.92에 이르렀다(12). AsymMirai는 Mirai의 사고 과정 중 일부를 이해하기 쉬운 형태로 구현함으로써, 임상적 활용성을 높이고 영상 판독 시 AI가 근거를 명확히 제시하는 데 기여할 것으로 기대된다. 다만, 저자들은 AsymMirai 역시 비대칭성 외의 요소나 데이터 편향 등의 한계는 여전히 존재한다고 언급하였다(12). 이 연구는 향후 영상 기반 위험 예측의 투

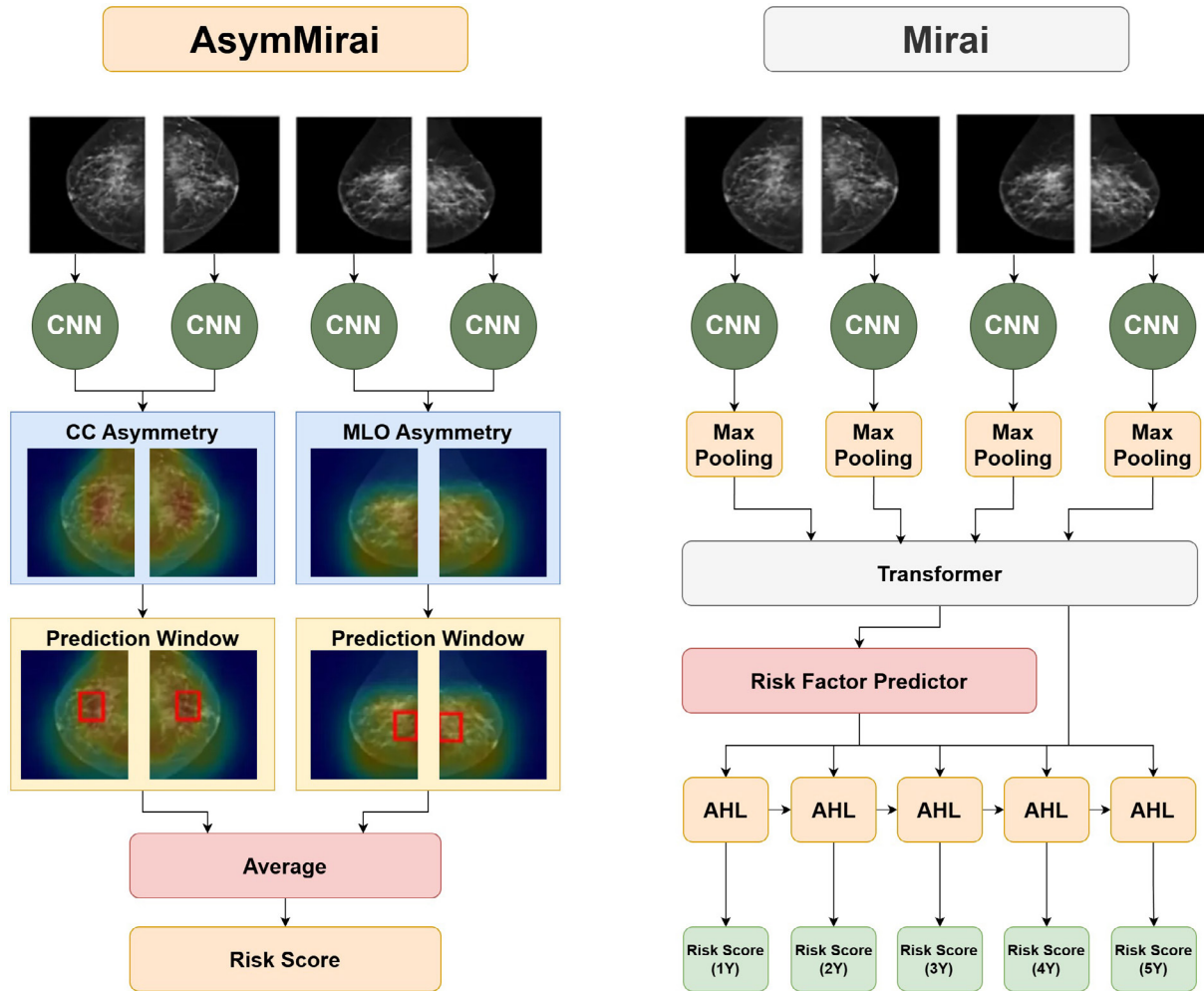


그림 1. AymMirai 및 Mirai 딥러닝 모델 구조도(참고문헌 12)

명성을 높이고 새로운 영상 표지자를 발굴하는 데 중요한 시사점을 제공한다.

3.2 CalcMirai 및 MassMirai: 영상 특징별 중요도 분석 모델

AsymMirai 연구로 좌우 비대칭성이 중요한 위험 예측 인자임이 밝혀졌지만, 보다 구체적인 영상 소견이 위험 예측에 미치는 영향에 대해 추가적인 연구가 필요하다. 다른 연구 그룹에서는 Mirai 모델의 특징 중에서 석회화(calcification) 및 종괴(mass)와 관련된 특징들을 중점적으로 분석하였다. Mirai가 추출한 512 차원의 이미지 특징 벡터 중 일부가 석회화 패턴과 강한 연관성을 보였다(13). 이 석회화 관련 특징 18개만을 사용하여 구성된 CalcMirai 모델(AUC 0.66)이 추가적인 학습 없이 전체

Mirai 모델(AUC 0.71)과 유사한 5년 예측 성능을 보였다. 이는 미세석회화가 Mirai 모델에서 위험 예측의 주요 영상 특징으로 활용되고 있음을 시사한다. 반면, 종괴와 관련된 특징만을 사용한 MassMirai의 경우 5년 예측 성능(AUC 0.57)이 크게 저하되었다(13). 이는 Mirai가 종괴보다는 미세석회화와 같은 조기 발견이 가능한 유방촬영술 이상 소견을 중심으로 유방암 위험도를 평가한다는 것을 시사한다.

Mirai의 예측 성능 배경에는 좌우 유방 간의 비대칭성과 미세석회화 두 가지 영상 특징이 핵심적인 역할을 하고 있는 것으로 보인다. 향후 딥러닝 기반 유방암 예측 모델을 개발할 때 이러한 특징들을 명시적으로 반영하거나, 추가적인 영상 소견을 탐색함으로써 모델의 성능과 해석력을 동시에 향상시킬 수 있을 것이다.

4. 설명 가능한 인공지능(XAI)의 개요

복잡한 딥러닝 모델의 의사 결정을 사람이 이해하기 쉽게 설명하려는 시도가 설명 가능한 AI(XAI, explainable AI)라는 분야로 발전했다. XAI 기법은 주로 모델이 예측을 내린 이유를 시각화하거나 입력 특징의 기여도를 정량화하여 설명을 제공한다. XAI는 크게 사후 설명(post-hoc explanation)과 내재적 해석 가능 모델(inherently interpretable model)로 나뉜다. AsymMirai처럼 구조적으로 해석이 가능하도록 설계된 모델은 내재적 해석 모델에 해당하며, 기존 복잡한 모델의 출력에 대해 추가 분석을 제공하는 방식은 사후 설명으로 분류된다.

4.1 시각적 설명 기법(CAM 계열)

의료영상 분야에서 가장 널리 활용되는 XAI 기법은 클래스 활성화 매핑(class activation mapping; 이하 CAM)과 이것의 변형이다(14). CAM은 원래 CNN 분류 모델에서 특정 클래스에 대한 활성화 정도를 시각적으로 나타내는 히트맵(heatmap)을 생성하는 기법이다. 이 방법은 마지막 합성곱층(convolutional layer)의 출력과 분류 가중치를 선형 결합하여, 원본 영상 위에 모델이 주목한 영역을 직관적으로 나타낸다.

이후 구조적 제약 없이 모든 CNN 기반 모델에 적용할 수 있도록 발전된 방법이 Grad-CAM(Gradient-weighted CAM)이다(15). Grad-CAM은 모델 출력에 대한 마지막 합성곱층 뉴런의 기여도를 그래디언트로 계산하여 특정 클래스의 중요도 맵을 생성한다(그림 2). 이를 통해 다양한 CNN 구조에서 직관적인 설명을 제공할 수 있으며, 의료영상 진단을 위한 여러 연구에서 널리 사용되고 있다.

최근에는 Grad-CAM을 더욱 발전시킨 Grad-CAM++ 및 픽셀 단위 중요도를 제공하는 적분 그래디언트(integrated gradients) 등이 제안되어 XAI의 정밀도가 더욱 향상되었다. 시각적 설명 방법은 정확한 병변 위치 강조 시 신뢰성을 높이지만, 부정확하거나 광범위한 강조는 오히려 신뢰를 저하시킬 수 있다(16, 17).

4.2 특징 중요도 및 예측 인자 기여도

XAI의 또 다른 주요 축은 모델의 입력 특징(feature)이 결과에 미치는 영향을 정량적으로 측정하는 것이다. 전통적인 위험 모델에서도 활용된 SHAP(SHapley Additive

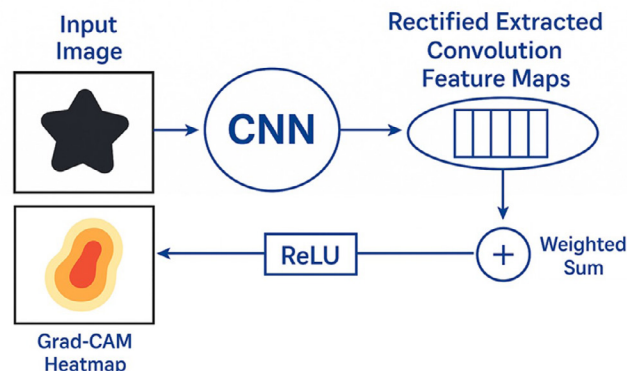


그림 2. Grad-CAM 시각화 기법의 작동 원리

exPlanations)는 최근 딥러닝 모델의 해석에도 널리 적용되고 있다(18). SHAP는 게임 이론에서 차용한 셰플리(Shapley) 값을 기반으로, 입력 변수 각각의 공헌도를 산출하여 복잡한 모델의 예측을 국소적으로 선형적 설명으로 제공한다. 영상의 경우, 픽셀 단위보다는 특정 영역이나 특징 벡터 단위로 SHAP 값을 계산하는 것이 일반적이다. 예를 들어 유방촬영 영상에서 특정 미세석회화 영역을 제거했을 때 모델의 예측 위험도가 크게 감소하면, 그 영역은 높은 양성 SHAP 값을 갖는다. 이 방법은 시각적 히트맵이 제공하는 직관적 정보에 더해 각 입력 특징의 방향성과 기여 크기까지 정량적으로 평가할 수 있어서 보다 세부적인 모델 해석이 가능하다. 다만, SHAP 방법은 많은 계산을 요구하여 복잡한 딥러닝 모델에 적용할 때 계산 비용이 높고, 상호 연관성이 큰 영상 특징에 적용할 때는 결과 해석에 주의가 필요하다.

5. 의료영상에서의 XAI 적용 사례

XAI 기법은 최근 다양한 의료영상 AI 연구에 널리 활용되며 유용성을 입증하고 있다. 뇌종양 MRI 분류 연구에서 Grad-CAM을 적용한 사례에 따르면, 생성된 히트맵이 병소 부위를 정확히 강조하여 모델이 학습한 영상 소견이 임상적 판단과 잘 부합함을 확인하였다(19). 또한 흉부 X선 영상으로 폐렴을 탐지하는 연구에서는 여러 특징 지도(saliency map) 기법을 비교했는데, Grad-CAM 계열 방법이 병변 위치를 가장 정확하게 강조하는 것으로 나타났다(18). 이러한 결과들은 XAI의 시각적 설명 기법이 의사의 진단을 보조하고 AI에 대한 신뢰를 높이는 데 중요한 역할을 한다는 점을 시사한다. 실제 임상에서도 AI가 설명한 히트맵을 제공하면, 영상의학과 의사가 AI 판단 근거를 파악하여 놓친 병변을 재확인하거나, AI 결과의 타당성을

평가할 수 있다. 이는 의료 AI 시스템의 신뢰성과 사용자 교육에도 크게 기여한다.

국내 의료 인공지능 연구에서도 유방촬영술에 XAI 기법을 적용한 대표적인 사례가 있다. 예를 들어, 국내 의료 AI 기업인 루닛(Lunit)의 Lunit INSIGHT MMG는 유방암의 조기 진단을 지원하기 위해 Grad-CAM 기반의 시각적 설명 기법을 도입하여, 모델이 주목한 병변 영역을 영상 위에 히트맵 형태로 제시한다. 이 기법은 특히 치밀 유방과 같이 판독이 어려운 영상에서 의사의 판단을 보조하는 데 유용하며, 실제 임상 판독 환경에서 AI의 투명성과 신뢰도를 향상시키는 데 기여하고 있다(20). 이러한 사례는 XAI가 단순한 기술적 해석 도구를 넘어, 실제 임상 활용 단계에서 핵심적 요소로 자리잡고 있음을 보여준다.

다만, 일부 연구에서는 Grad-CAM 등의 XAI 기법이 항상 정확한 예측을 보장하는 것은 아니라고 지적한다. 데이터의 편향된 특징을 학습한 경우에도 히트맵이 부정확한 위치를 강조할 수 있어 주의가 필요하다(21). 따라서 시각적 설명 결과를 절대적으로 신뢰하기보다는, 의사의 판단과 함께 보조적 도구로 활용하는 것이 바람직하다.

6. Mirai 모델에서의 XAI 적용

일반적인 질병 분류 모델과는 다르게 Mirai와 같은 위험도 예측 모델에 XAI 기법을 적용하는 것은 어렵다. 질병 분류는 병변 위치를 명확히 시각화하는 반면, 위험 예측은 미래의 불확실한 사건(암 발생 확률)을 평가하기 때문에 영상 전반에 걸친 미묘한 패턴과 복합적인 요인이 복잡하게 작용한다. Mirai 개발팀도 Grad-CAM과 같은 기존 시각화 기법을 시도했으나, Mirai 모델이 다중 뷰 입력 및 트랜스포머 기반의 복잡한 구조를 갖고 있어 명확한 설명이 어렵다고 하였다.

이러한 한계를 극복하기 위해 개발된 AsymMirai 모델은 주요 예측 근거로 좌우 유방의 국소적 비대칭성을 측정하고 이를 시각화하여 제시한다(12). 유방암 위험 예측 모델에 추가적으로 SHAP 분석 등을 적용하면, 좌우 유방 영상의 상대적 위험 기여도나 연속 촬영 시 나타난 미세한 영상 변화를 정량적으로 평가할 수 있을 것으로 기대한다.

결론

딥러닝 기반 영상 분석 기술은 유방암 위험 예측 분야에 새로운 지평을 열고 있다. Mirai와 같은 영상 기반 예측 모델은 임상 정보 없이도 유방촬영 영상의 미세한 패턴을

포착하여, 향후 수년 내 유방암 발생 가능성을 예측할 수 있다. Mirai는 다기관 검증을 통해 뛰어난 성능과 안정성을 입증하였으며, 이를 기반으로 개인 맞춤형 검진(risk-stratified screening)에도 도움을 줄 수 있다(22). 하지만 이러한 딥러닝 모델의 의료 현장 수용성을 높이기 위해서는 모델의 예측 근거를 투명하게 제공하는 설명 가능한 인공지능(XAI) 기술이 필수적이다(23). 최근의 AsymMirai와 같은 연구는 영상 기반 위험 예측 모델의 해석 가능성을 높이는 새로운 접근법을 제시했다. 결론적으로 설명 가능한 AI를 통한 예측 근거 제시의 의료진과 환자 간 소통을 강화하고, 정밀의료 시대의 유방암 예방 및 조기검진 전략의 발전을 이끌 중요한 열쇠가 될 것이다.

참고 문헌

1. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81(24):1879-1886.
2. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med 2004;23(7):1111-1130.
3. Brentnall AR, Harkness EF, Astley SM, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. Breast Cancer Res 2015;17(1):147.
4. Colditz GA, Bohlke K. Priorities for the primary prevention of breast cancer. CA Cancer J Clin 2014;64(3):186-194.
5. Tice JA, Cummings SR, Smith-Bindman R, et al. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Ann Intern Med 2008;148(5):337-347.
6. Yala A, Lehman C, Schuster T, et al. A deep learning mammography-based model for improved breast cancer risk prediction. Radiology 2019;292(1):60-66.
7. Portnoi T, Yala A, Schuster T, et al. Deep learning model to assess cancer risk on the basis of a breast MR image alone. AJR Am J Roentgenol 2019;213(1):227-233.
8. Yala A, Mikhael PG, Strand F, et al. Multi-institutional validation of a mammography-based breast cancer risk model. J Clin Oncol 2022;40(16):1732-1740.
9. Yala A, Mikhael PG, Strand F, et al. Toward robust

- mammography-based models for breast cancer risk. *Sci Transl Med* 2021;13(578):eaba4373.
10. Damiani C, Kalliatakis G, Sreenivas M, et al. Evaluation of an AI model to assess future breast cancer risk. *Radiology* 2023;375(5):e222679.
11. Avendano D, Marino MA, Bosques-Palomo BA, et al. Validation of the Mirai model for predicting breast cancer risk in Mexican women. *Insights Imaging* 2024;15(1):244.
12. Donnelly J, Moffett L, Barnett AJ, et al. AsymMirai: interpretable mammography-based deep learning model for 1-5-year breast cancer risk prediction. *Radiology* 2024;310(3):e232780.
13. Wang YK, Klanecek Z, Wagner T, et al. Longitudinal analysis of micro-calcifications features for breast cancer risk prediction with the Mirai model. 17th International Workshop on Breast Imaging (IWBI 2024), SPIE Proceedings Vol. 13174.
14. Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016.
15. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision* 2020;128:336-359.
16. Chattopadhyay A, Sarkar A, Howlader P, et al. Grad-CAM+: generalized gradient-based visual explanations for deep convolutional networks. *IEEE Winter Conference on Applications of Computer Vision (WACV)* 2018.
17. Chien J-C, Lee J-D, Hu C-S, et al. The usefulness of gradient-weighted CAM in assisting medical diagnoses. *Appl Sci* 2022;12(15):7748.
18. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. 31st International Conference on Neural Information Processing Systems (NIPS 2017).
19. Zhang Y, Hong D, McClement D, et al. Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. *J Neurosci Methods* 2021;353:109098.
20. 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(3):e138-e148.
21. Saporta A, Gui X, Agrawal A, et al. Benchmarking saliency methods for chest X-ray interpretation. *Nature Machine Intelligence* 2022;4:867-878.
22. Bae MS. Mammography-based deep learning for breast cancer risk assessment for supplemental MRI screening. *Radiology* 2023;308(3):e232226.
23. 도신희. 의료 영상 분석을 위한 설명 가능하고 안전한 인공지능. *J Korean Soc Radiol* 2024;85(5):834-847.