

대한유방영상의학회/대한유방검진학회 지원 다기관 연구 보고서

지원프로그램	다기관 연구		연구관리번호		KSBI&KSFBS-2019-02
연구 과제명	국 문	검진 유방 초음파의 위양성 병변 저감을 위한 인공 지능 컴퓨터 보조진단 (S-Detect) 적용의 모델 구축을 위한 다기관 연구			
	영 문	Multicenter study for modeling artificial intelligence based computer-aided diagnosis (S-detect) in application of screening ultrasound detected lesion to reduce false positive biopsy			
연구책임자	성 명	장정민 (張禎珉)			
	소 속	서울대학교병원 영상학과			
	주 소	서울시 종로구 대학로 101			
	직 위	교수	전공	유방영상의학	
연구비	2,000 만원	연구기간	2019 년도 1 월 1일 부터 2019 년도 12 월 31일 까지 (1 년)		
공동 연구자 및 참여기관	김수연 (서울대학교병원) 김은경 (연세의대 용인세브란스병원) 윤정현 (연세의대 신촌세브란스병원) 한부경 (성균관의대 삼성서울병원) 최지수 (성균관의대 삼성서울병원)				

본 연구과제는 대한유방영상의학회/대한유방검진의학회의 지원 (KSBI&KSFBS-2019-02, 서울대병원 위탁과제 06-2019-0540)을 받아 수행되었습니다.

본 결과보고서에 사용된 표와 그림은 연구결과 출판물 (Sci Rep. 2021;11(1): 395) 에 사용된 표와 그림을 발췌하였음을 밝힙니다.

I. 연구 결과 요약문

- 치밀 유방검진을 위해 시행하는 초음파 검사에서 우연히 발견되는 많은 병변들은 그 크기가 작고, 전형적인 악성 소견이 결여되어 감별진단이 어려울 때가 많음.
- 딥러닝 기반의 컴퓨터 보조 진단 (Deep learning-based computer-aided diagnosis, DL-CAD)을 통해 획득된 병변의 형태학적 특징과 임상적인 특징을 종합하여 검진 초음파에 적용할 경우, 초음파에서 발견된 병변의 감별진단에 도움을 받을 수 있을 것으로 가설하에 연구를 시행함.
- 검진목적의 유방 초음파 발견 유방 종괴에 대한 DL-CAD 데이터가 있는 여성의 영상을 서울대학교병원, 신촌세브란스병원, 삼성서울병원에서 수집하여 후향적으로 분석하였음. 수집한 데이터 중 신촌세브란스병원, 삼성서울병원에서 획득된 데이터(2015년도 299명의 영상자료)를 모델 개발군 자료로 활용되었음. 병변의 형태학적인 특징에 대한 정량적인 정보는 DL-CAD 소프트웨어 프로그램을 통해 획득되었고, 각 케이스의 임상자료도 함께 수집되었음.
- 모델 개발군에서 획득한 정량적 영상 정보 및 임상 정보를 이용하여 다변량 로지스틱 회귀 분석을 시행하여 유의한 인자를 추출하여 이를 이용하여 진단 노모그램을 개발하였음.
- 개발한 진단 노모그램의 검증을 위하여 서울대학교병원에서 획득한 164명의 영상 (2018년부터 2019년)을 이용하여 검증을 진행함.
- DL-CAD 소프트웨어 분석 결과 병변에서 추출한 형태학적 특징 중에 높은 불규칙한 모양 점수 ($P = .018$)와 낮은 평행 방향 점수 ($P = .007$)가 병변의 악성도와 연관되었음.
- DL-CAD 소프트웨어 데이터의 유의한 형태학적 특징 항목들과 판독 의사의 Breast Imaging Reporting and Data Systems (BI-RADS) 최종 범주 ($P = .014$), 환자 나이 ($P < .001$)를 종합한 노모그램은 개발 및 검증군 모두 좋은 식별능력을 보였음.
- 판독 의사의 BI-RADS 최종 범주와 노모그램에서 114점을 기준으로 하여 비교할 때, 민감도는 개발군 (98% vs. 95%, $P = .317$) 및 검증군 (각각 100%)에서 모두 변화를 보이지 않았고, 위양성율을 두 군에서 모두 낮출 수 있었음 [68% vs. 31%, $P < .001$ (개발군); 97% vs. 45% $P < .001$ (검증군)].

II. 연구내용 및 결과

1. 연구 과제의 개요

1) 치밀 유방에서의 유방암 검진과 검진 초음파의 한계

- 유방 촬영술은 유방암의 조기 발견과 유방암 관련 사망률 및 이환율을 줄이는 것으로 알려져 있음. 그러나 유방 촬영술은 치밀 유방을 가진 여성에서는 민감도가 떨어짐. 따라서 치밀 유방에서의 유방암 검진을 위해 보조적인 영상기법의 성적에 대한 연구들이 활발하게 진행되었음.
- 선별 유방 초음파 검사는 유방촬영술상 정상 소견을 여성에서 가장 널리 사용되는 영상임. 발견되는 암의 수는 암의 유병률에 따라 다르지만 초음파검사는 일반적으로 1,000건의 검사 당 평균 3-4개의 유방 촬영에서 보이지 않는 암을 찾아내는 것으로 알려져 있음. 초음파에서 발견된 대부분의 암은 크기가 작고 림프절 전이가 없는 조기 침윤성 암으로, 이런 암을 조기에 발견하는 것은 사망률을 낮추는 데 도움이 될 수 있을 것으로 기대됨.
- 유방 초음파 검진은 여러 가지 이점에도 불구하고, 특이도와 양성예측도가 낮고 위양성을 증가시키는 문제가 있음. 유방 영상 보고 및 데이터 시스템 (Breast Imaging Reporting and Data System, BI-RADS)의 활용은 유방 초음파 판독 표준화에 기여한 바가 크고, 그로 인하여 양성 및 악성 유방 병변을 구별하는데 도움이 되고 있으나, 작은 암을 하나라도 더 찾고자 하는 노력은 위양성을 증가시킬 수 밖에 없으며, 검진 환경에서 발견되는 암들은 그 크기가 작고, 악성 특징이 결여된 경우가 많아 진단이 더 어려울 수 있음. 따라서, 초음파 검진 중에 발견된 작은 병변의 악성 및 양성 감별은 어려우며, 시행자에 따라 다른 결과가 나오기도 함.

2) 유방 초음파에서의 인공지능의 활용

- 최근 인공지능, 특히 딥러닝 알고리즘은 이미지 인식 분야에서 탁월한 성능으로 주목을 받고 있음.
- 인공지능은 유방 종괴에 대한 자동 정량 평가가 가능하며, 유방 초음파 이외의 다른 여러 영상기기에서 사용될 수 있음. 딥러닝 기반 컴퓨터 보조 진단 (Deep learning-based computer-aided diagnosis; DL-CAD) 소프트웨어가 개발되어 상용화되었고, 특히 경

험이 적은 판독자의 유방암 진단 특이도를 높여주는 결과를 보여주었음.

- 기존의 유방 영상과 관련된 인공지능, 컴퓨터 보조 진단 연구들은 다양한 조건의 환자들을 포함하였고, 특히 초음파에서 유방 종괴에 대한 컴퓨터 보조 진단 성능이 많이 연구되지 않았음. 초음파 검진을 통해 발견된 유방 종괴가 대부분 그 크기가 작고 덜 분명한 형태학적 특성을 가진다는 점을 고려할 때, DL-CAD 소프트웨어에서 획득한 정량적 정보는 영상의학과 의사가 간과할 수 있는 정보를 제공할 수 있어 병변 감별과 환자 진료에 유용할 수 있음.
- DL-CAD 소프트웨어의 하나로 S-Detect은 삼성메디슨에서 개발하여 최근 상용화된 프로그램으로, 인공지능을 이용한 컴퓨터 보조 진단임. 이 프로그램을 유방 종괴에 적용하면, BI-RADS lexicon에 따른 평가와 최종 진단이 자동적으로 평가되어 화면에 표시됨. 따라서 S-detect은 보조적인 진단도구로 이용될 수 있고, 특히 경험이 적은 사람에서 유용하게 쓰일 수 있을 것으로 기대됨. 국내 단일 기관 연구에서 유방 종괴에 대한 S-Detect의 임상적용을 본 연구들이 있었음. 이 연구들의 결과를 종합하면, S-Detect을 단독으로 또는 B-mode 초음파와 병합으로 적용하였을 때, 진단의 특이도 (specificity)가 향상되었음. 특히, 경험이 적은 1년차 전공의에서는 특이도 뿐만 아니라 민감도 (sensitivity)도 향상되는 결과를 보임. 갑상선 종괴에서도 S-detect의 성적은 경험이 많은 영상의학과 의사의 성적과 비슷하였고, 영상의학과 의사 단독성과 비교시 S-Detect을 적용했을 때 민감도가 유의하게 향상되었음.

3) 연구 목적

- 본 연구에서는 초음파에서 발견된 암과 양성 병변의 감별진단에 있어 도움이 되는 형태학적 정보를 DL-CAD 소프트웨어를 사용하여 발굴하고, 이러한 DL-CAD 정량 정보와 임상정보를 통합하는 진단 모델을 개발하여 이를 이용한 진단 성능을 알아보고 인공지능 모델을 이용한 위양성 감소효과를 알아보고자 하였음.

2. 연구수행 내용 및 결과

1) 연구 수행 대상 및 방법

- 본 연구는 삼성메디슨(주) Samsung Medison Co. (성남, 대한민국)에서 개발한 DL-CAD 소프트웨어 (S-Detect; 삼성메디슨(주), 성남, 대한민국)를 활용하여 병변의 데이터를 추출하였다. 본 연구는 후향적인 연구로 각 기관의 임상 윤리 심의 위원회의 승인을 받았으며, 동의서는 면제되었음.
- 환자 모델 개발군을 구축하기 위하여, 연세대학교 신촌세브란스 병원 및 성균관대학교 삼성서울병원에서 2015년 1월부터 12월까지 얻은 초음파 영상을 후향적으로 분석하였음. 이들 중 초음파에서 발견된 병변으로 초음파 유도 하 조직검사를 시행한 여성들만 포함하였음. 유방촬영에서 이상 소견이 보이거나 증상이 있는 여성, 유방암 병력이 있는 여성은 모두 제외하였고, 최종적으로 총 299명 여성 (평균 연령 45세, 19-81세) 이 포함되었음.
- 외부 검증을 위한 모델 검증군의 자료는 2018년 2월부터 2019년 6월까지 서울대병원에서 획득된 DL-CAD 영상이 있는 164명의 여성 (평균연령, 45세; 23-74세)에게서 수집되었음. 임상 병리학적 정보, 나이, 조직검사 및 수술 결과, 초음파 판독 결과등이 수집되었음. 전체 환자중 133명에서 수술이 시행되었음. 조직검사후 양성 판정을 받은 경우, 평균 추적검사기간은 12개월 (범위: 6-23개월)이었고, 추적검사 기간 중 크기 모양변화를 보이는 병변은 없었음 (그림 1).
- 모든 초음파 영상의 획득은 최신 초음파 장비를 이용하여 (RS 80A with Prestige, RS85A Prestige; 삼성메디슨(주), 성남, 대한민국) 초음파 유도 하 조직검사 전에 시행되었음. 6명의 유방 전문의 중의 한 명이 초음파 검사를 시행 후 BI-RADS 판정에 따라 병변의 최종 범주를 제시하였음. 이후, 병변의 대표영상을 선택하여 저장한 후, DL-CAD 소프트웨어를 적용하였고, 병변의 모양특성과 최종범주에 대한 정량 정보를 추출하였음.
- 정량정보는 종괴의 모양, 방향, 변연부, 후방음영, 에코 특성을 포함하여 0에서 1 사이의 숫자로 표현되었음 (그림 2). 조직검사의 시행여부는 DL-CAD 결과와 무관하게 결정되었음.
- 연구의 결과 분석을 위하여 DL-CAD 소프트웨어에서 추출한 병변 모양의 정량 정보, 환자의 나이, 병변의 크기, 판독의사가 평가한 병변의 BI-RADS 최종 범주, 그

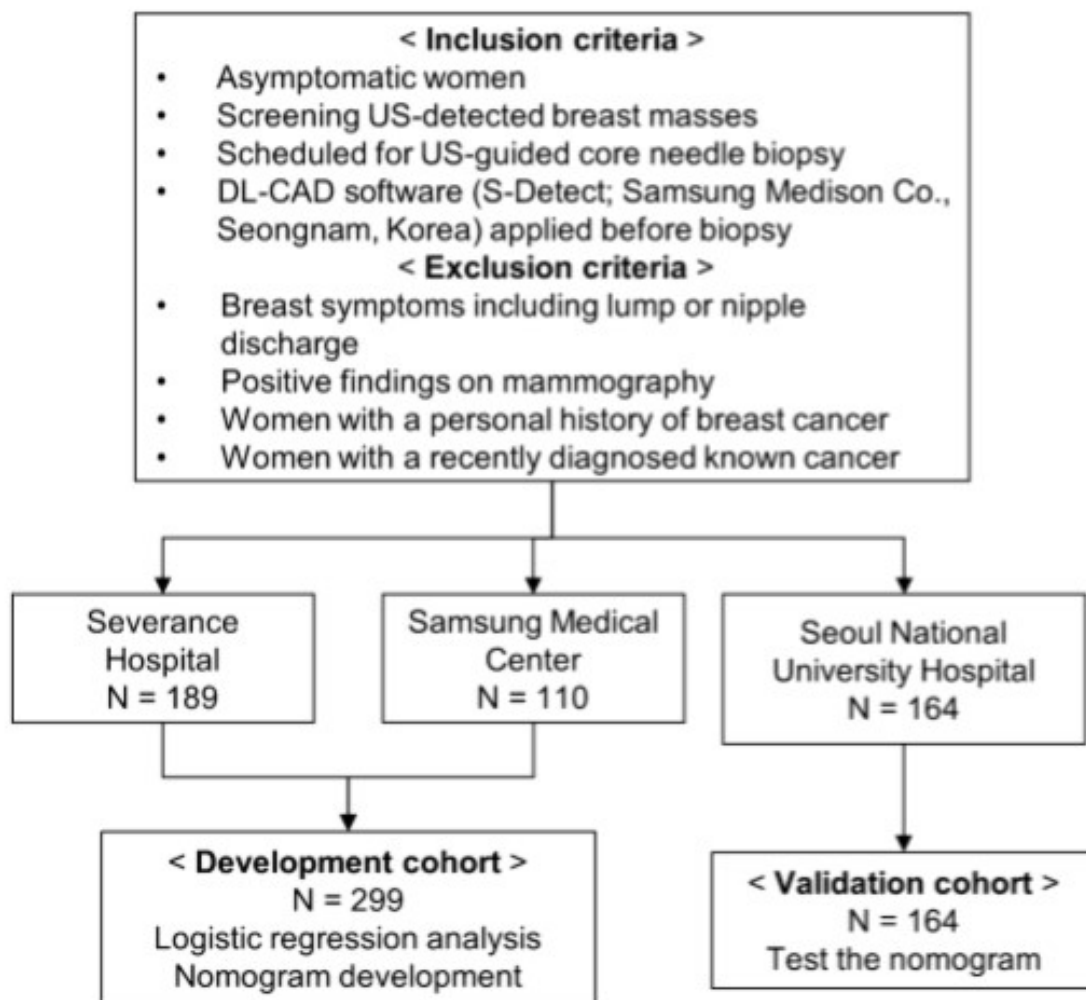


그림 1. 환자 모집 및 진단 모델 개발군과 모델 검증군의 구성

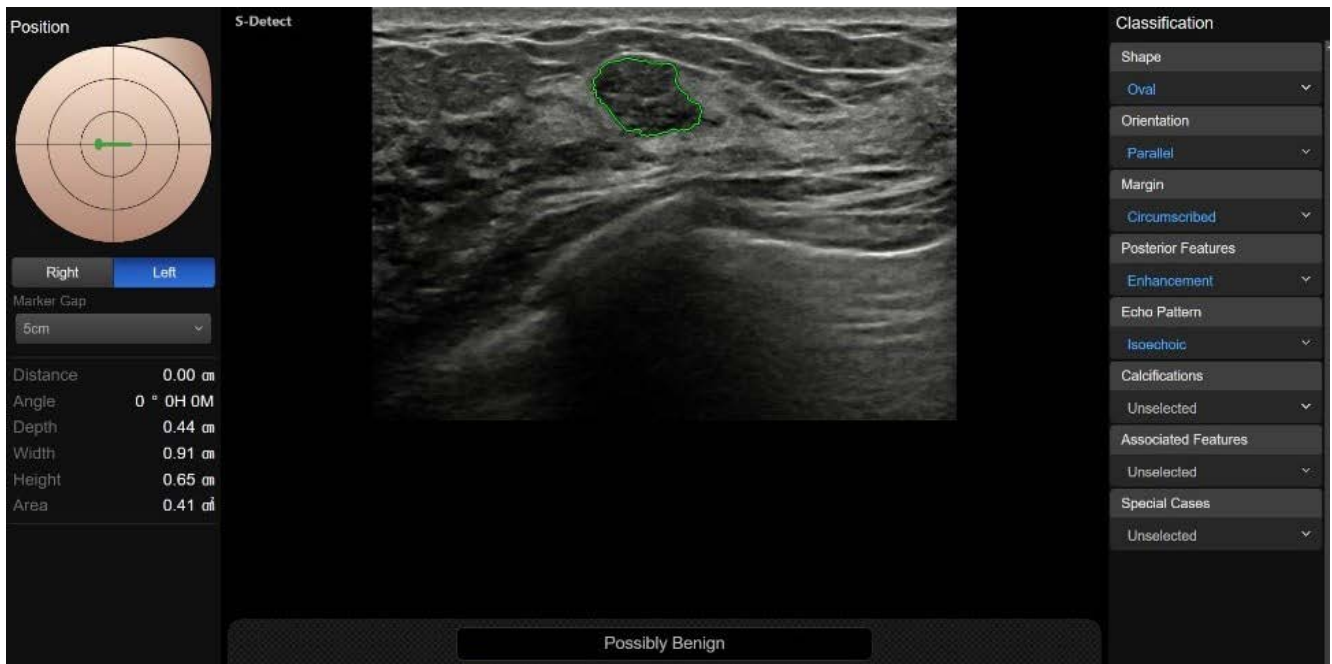
리고 병리 결과 정보가 수집되었음.

- 진단모델 개발군과 검증군의 환자요인 비교, 모델 개발군내의 악성병변 및 양성병변의 특징비교를 위하여 Chi-squared test, Wilcoxon rank sum test 와 independent t-test가 사용되었음. 모델 개발군에서 암 진단의 유의한 인자 발굴을 위하여, 단변량 및 다변량 로지스틱 회귀분석이 시행되었고, 단변량 분석에서 P value 0.1 미만을 보인 변수들만 다변량 분석에 포함되었음. 다변량 로지스틱 회귀분석을 이용하여 노모그램을 개발하였고, 노모그램의 성능은 ROC (receiver operating characteristic) 커브의 AUC(area under the receiver operating characteristic curve) 값을 계산하여 산출하였음.
- 모델 개발군에서 노모그램 적용을 가정하였을 때와 영상의학과 의사의 단독 판독 (BI-RADS 최종 범주 4A

이상 조직검사 권유)과 비교할 때, 민감도의 감소 없이 (최소 95% 이상의 민감도 유지) 특이도를 최대한 높이고, 위양성율을 줄이는 기준 값을 확인하였음. 모델 검증군에 대하여 해당 노모그램 기준을 적용하였고, 적용전의 영상의학과 의사의 단독 판독과 노모그램 적용 후의 민감도, 특이도, 위양성율을 비교하여 진단 모델의 성능을 평가하였음. 통계 분석을 위하여 SAS 소프트웨어 (version 9.2; SAS Institute, Cary, NC)와 R package (version 3.6.1)가 사용되었음.

2) 연구 결과

- 최종 463명 (평균 연령 45세, 범위 19-81)에서 초음파 영상과 DL-CAD 소프트웨어 적용 결과를 획득하였음. 모델 개발군과 검증군의 특성은 표 1에 있음.



	Shape		Orientation		Margin		Posterior feature		Echogenicity	
	round	0.053037	parallel	0.990681	circumscribed	0.913474	no posterior features	0.597901	anechoic	0.003545
	oval	0.88724	not parallel	0.009319	indistinct	0.01014	enhancement	0.401953	hyperechoic	0.005622
	irregular	0.059723			spiculated	0.000001	shadowing	0.000017	complex	0.016031
					angular	0.000004	combined	0.000129	hypoechoic	0.250737
					microlobulated	0.076381			isoechoic	0.715054
									heterogeneous	0.009011
Total	1		1		1		1		1	
Output	oval		parallel		circumscribed		no posterior features		isoechoic	

그림 2. DL-CAD 소프트웨어를 이용한 병변의 정량 특징 추출

- 진단 모델 개발군과 검증군에서 환자의 나이 ($P = .937$)와 유방 병변의 크기 ($P = .044$)는 비등하였음. 모델 개발군에서 모델 검증군에 비해 BI-RADS 최종 범주 3의 빈도 (28.1% [84 of 299] vs. 2.4% [4 of 164], $P < .001$) 및 암 유병율이 유의하게 높았음 (14.4% vs. 5.5%, $P = .004$). 모델 개발군에서 양성병변은 256개 (85.6%), 악성 병변은 43개 (14.4%)였고 모델 검증군에서는, 양성 병변은 155개 (94.5%)였고, 암은 9개 (5.5%) 였음.
- 진단 모델 개발군의 양성 및 악성 병변의 특징을 단변량 로지스틱 회귀 분석으로 분석한 결과 (표 2), 악성 병변을 가진 여성이 양성 병변을 가진 여성보다 나이가 많고 (평균연령, 54 vs. 44 years, $P < .001$), BI-RADS 4A 이상의 판정을 받은 경우가 많았음. DL-CAD 소프트웨어를 이용한 병변 모양 평가에 있어서

는 악성 병변이 양성 병변보다 낮은 oval shape 점수 (중위값, 0.15 vs. 0.77, $P < .001$), 높은 irregular shape 점수 (0.82 vs. 0.12, $P < .001$), 낮은 parallel orientation 점수 (0.72 vs. 0.99, $P < .001$), 높은 not parallel orientation 점수 (0.28 vs. 0.01, $P < .001$), 낮은 circumscribed margin 점수 (0.11 vs. 0.98, $P < .001$), 높은 spiculated margin 점수 (0 [interquartile ranges, 0-0.01] vs. 0 [0-0], $P = .037$), 높은 microlobulated margin 점수 (0.16 vs. 0.01, $P < .001$), 낮은 posterior enhancement 점수 (0.14 vs. 0.38, $P < .001$), 높은 posterior shadowing 점수 (0.01 vs. 0, $P = .011$)를 보였음.

- 다변량 로지스틱 회귀 분석을 시행했을 때, 최종적으로 4개의 변수가 암 진단과 연관성을 보였고, 이들은 BI-RADS 4A 이상 (odds ratio [OR], 14.7 [95%

표 1. 모델 개발군과 검증군의 특징 비교

Characteristic		Development cohort (n = 299)	Validation cohort (n = 164)	P value
Age (y)		45 ± 12 (19-81)	45 ± 10 (23-74)	.937
Size on US (cm)		1.1 ± 0.6 (0.4-3.2)	1.1 ± 0.5 (0.3-3.4)	.444
Radiologist's BI-RADS assessment	3	84 (28.1)	4 (2.4)	<.001
	4A	174 (58.2)	146 (89.0)	
	4B	20 (6.7)	11 (6.7)	
	4C	14 (4.7)	3 (1.8)	
	5	7 (2.3)	0 (0)	
Pathology	Benign	256 (85.6)	155 (94.5)	.004
	Fibroadenoma	128 (50.0)	57 (36.8)	
	Intraductal papilloma	28 (10.9)	14 (9.0)	
	Fibroadenomatoid change	28 (10.9)	13 (8.4)	
	Fibrocystic change	21 (8.2)	16 (10.3)	
	Stromal fibrosis	11 (4.3)	7 (4.5)	
	Sclerosing adenosis	8 (3.1)	14 (9.0)	
	Duct ectasia	6 (2.3)	5 (3.2)	
	Benign phyllodes tumor	4 (1.6)	6 (3.9)	
	Radial scar	4 (1.6)	1 (0.6)	
	Atypical ductal hyperplasia	3 (1.2)	5 (3.2)	
	Chronic inflammation	3 (1.2)	0 (0)	
	Galactocele	3 (1.2)	0 (0)	
	Fat necrosis	2 (0.8)	1 (0.6)	
	Usual ductal hyperplasia	1 (0.4)	6 (3.9)	
	Columnar cell change	0 (0)	7 (4.5)	
	Others*	6 (2.3)	3 (1.9)	
	Malignant	43 (14.4)	9 (5.5)	
	Invasive ductal carcinoma	33 (76.7)	7 (77.8)	
	Ductal carcinoma in situ	7 (16.3)	2 (22.2)	
	Invasive lobular carcinoma	1 (2.3)	0 (0)	
	Mucinous carcinoma	1 (2.3)	0 (0)	
	Adenoid cystic carcinoma	1 (2.3)	0 (0)	

Note.- Data for age and size are mean ± standard deviations (ranges). Data for BI-RADS final assessment and pathology are numbers (percentages). * 2 lobular carcinoma in situ, 2 papillary apocrine metaplasia, 1 mucocoele-like lesion, and 1 ductal epithelial hyperplasia in the development cohort, 1 glanulomatous lobular mastitis, 1 adenomyoepitheloma, 1 cholesterol granuloma in the validation cohort.

confidence interval (CI): 1.7, 125.6], $P = .014$), 높은 연령 (OR, 1.1 [95% CI: 1.0, 1.1], $P < .001$), DL-CAD 소프트웨어의 높은 irregular shape 점수 (OR, 5.1 [95% CI: 1.3, 19.4], $P = .018$), and 낮은 parallel orientation 점수 (OR, 0.1 [95% CI: 0, 0.6], $P = .007$) 였음

- 상기 4개의 변수를 포함한 노모그램은 그림 3 과 같음.
- 진단모델 개발군에서 영상의학과 판독을 기준으로 한

진단의 AUC 값은 0.65 (95% CI: 0.61, 0.69) 였던 반면 노모그램을 이용한 AUC 값은 0.89 (95% CI: 0.84, 0.93) 였음. 모델 검증군에서도 노모그램을 이용한 AUC 값은 0.87 (95% CI: 0.79, 0.95)로 영상의학과 의사의 BI-RADS 판정으로 평가한 성적 0.51 (95% CI: 0.50, 0.53) 보다 높았음 ($P < .001$) (그림 4).

노모그램을 이용한 조직검사 결정을 위한 적합한 점수

표 2. 모델개발군에서 양성 및 악성 병변의 임상 및 영상 특징 비교

Characteristic		Benign (n = 256)	Malignant (n = 43)	Univariable Odds ratio (95% CI)	Univariable P value
Age (y)		44 ± 11 (19-78)	54 ± 13 (22-81)	1.1 (1.0, 1.1)	<.001
Size on US (cm)		1.1 ± 0.5 (0.4-3.2)	1.3 ± 0.9 (0.4-3.2)	1.5 (0.9, 2.4)	.124
Radiologist's BI-RADS assessment	3	83 (32.4)	1 (2.3)	Reference	
	≥ 4A	173 (67.6)	42 (97.7)	20.2 (2.7, 148.9)	.003
Quantitative morphology scores from DL-CAD software*					
Characteristic	Descriptor				
Shape	Round	0.01 (0, 0.05)	0.01 (0, 0.08)	0.6 (0, 11.3)	.734
	Oval	0.77 (0.36, 0.93)	0.15 (0.05, 0.51)	0.04 (0.01, 0.14)	<.001
	Irregular	0.12 (0.03, 0.49)	0.82 (0.44, 0.92)	18.6 (6.3, 54.8)	<.001
Orientation	Parallel	0.99 (0.95, 0.99)	0.72 (0.24, 0.96)	0.03 (0.01, 0.1)	<.001
	Not parallel	0.01 (0, 0.05)	0.28 (0.05, 0.76)	30.9 (9.4, 101.0)	<.001
Margin	Circumscribed	0.98 (0.41, 0.99)	0.11 (0, 0.84)	0.2 (0.1, 0.4)	<.001
	Indistinct	0 (0, 0.05)	0.02 (0, 0.46)	2.3 (0.8, 6.6)	.107
	Spiculated	0 (0, 0)	0 (0, 0.01)	9.6 (1.2, 80.9)	.037
	Angular	0 (0, 0)	0 (0, 0)	1.2 (0.04, 38.9)	.902
	Microlobulated	0.01 (0, 0.07)	0.16 (0.01, 0.72)	6.5 (2.4, 17.5)	<.001
Posterior features	No	0.44 (0.18, 0.71)	0.58 (0.20, 0.78)	2.1 (0.7, 6.3)	.195
	Enhancement	0.38 (0.12, 0.72)	0.14 (0.01, 0.30)	0.1 (0.03, 0.4)	<.001
	Shadowing	0 (0, 0.01)	0.01 (0, 0.32)	4.1 (1.4, 12.3)	.011
	Combined	0 (0, 0)	0 (0, 0.02)	2.0 (0.3, 13.4)	.465
Echo pattern	Anechoic	0 (0, 0)	0 (0, 0)	1.1 (0.9, 1.2)	.281
	Hyperechoic	0 (0, 0)	0 (0, 0)	3.9 (0, 836.3)	.618
	Complex	0 (0, 0)	0 (0, 0)	0.4 (0, 37.2)	.671
	Hypoechoic	0.84 (0.19, 0.98)	0.91 (0.28, 0.99)	1.7 (0.7, 4.2)	.292
	Isoechoic	0.05 (0, 0.46)	0.01 (0, 0.29)	0.6 (0.2, 1.7)	.312
	Heterogeneous	0 (0, 0)	0 (0, 0.03)	2.7 (0.6, 12.9)	.213

Note.- CI = confidence interval. Data for age and size are mean ± standard deviations (ranges). Data for BI-RADS final assessment and pathology are numbers (percentages). * Quantitative morphology scores are median (25th percentile, 75th percentile).

표 3. 악성 병변 진단을 위해 진단 노모그램에 포함된 최종 변수

	Odds ratio (95% CI)	P value
Age	1.1 (1.0, 1.1)	<.001
Radiologist's BI-RADS assessment ≥ 4A	14.7 (1.7, 125.6)	.014
Irregular shape score	5.1 (1.3, 19.4)	.018
Parallel orientation score	0.1 (0, 0.6)	.007

는 민감도를 95% 이상 유지하면서 특이도를 최적화할 수 있는 조건으로 114점으로 결정되었음. 노모그램을 적용함으로써, 모델 개발군에서 위양성은 68%에서 31%으로, 모델 검증군에서 위양성은 97%에서 45%으로 감소하였음.

조직검사율은 모델 개발군에서 72%에서 41%, 모델 검증군에서 98%에서 48%으로 감소하였음. 민감도는 모델 개발군 및 검증군에서 모두 유의한 변화를 보이지 않았음.

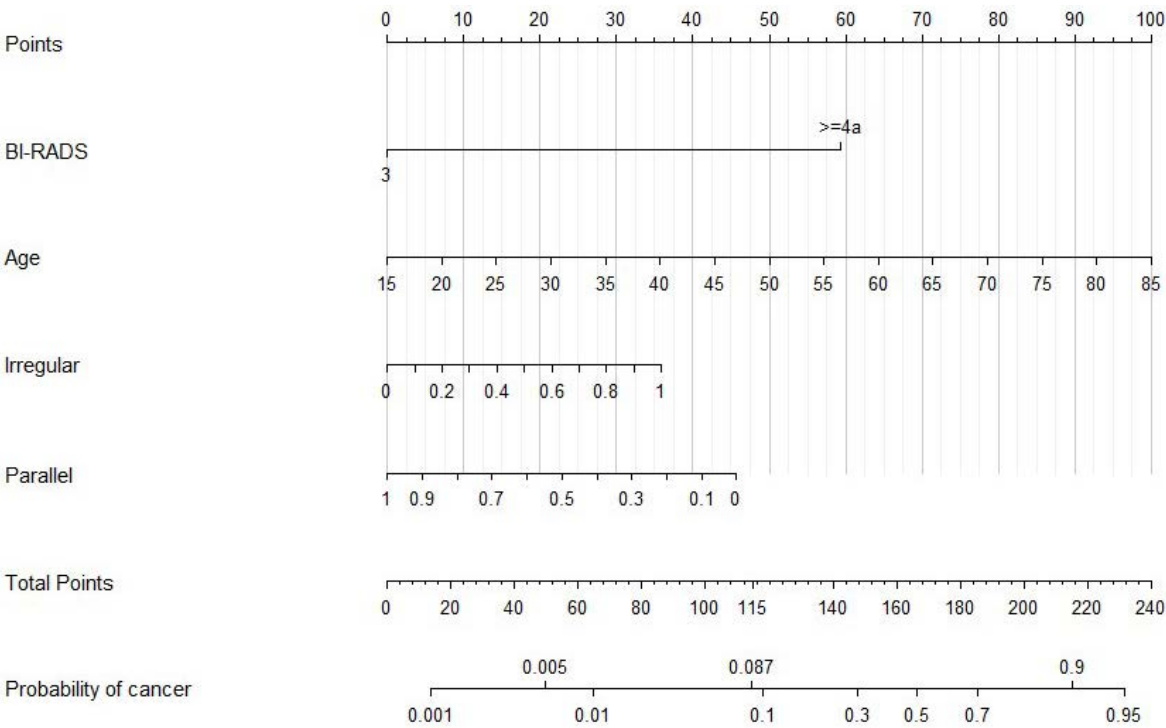


그림 3. 개발된 진단 노모그램

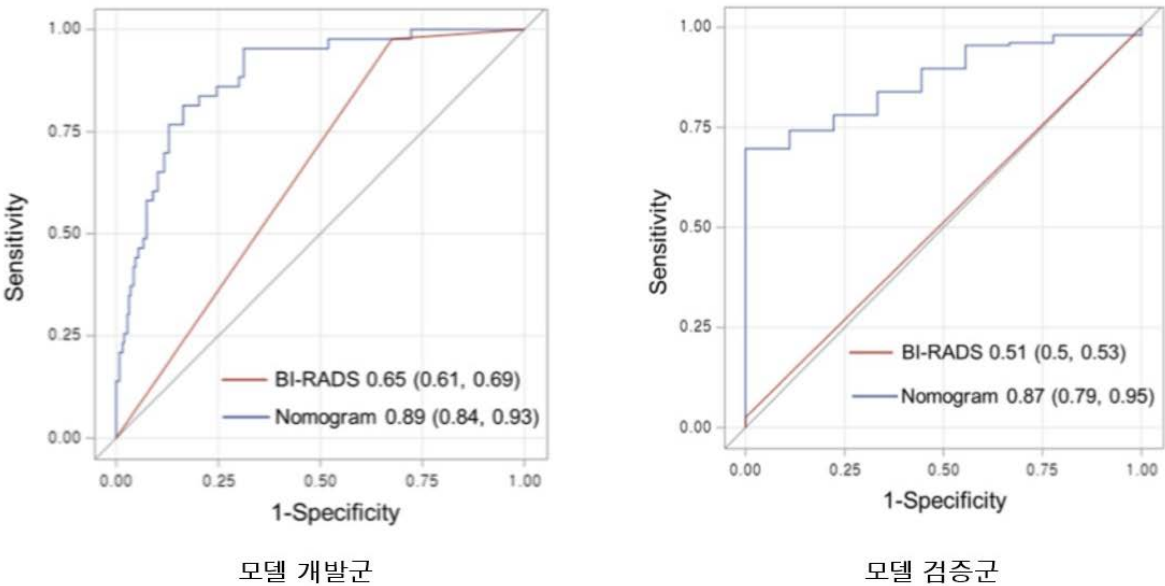
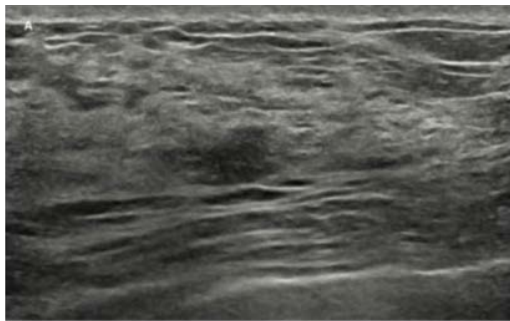


그림 4. 진단 모델 개발군과 검증군에서의 BI-RADS 단독 판정과 노모그램 적용의 성적

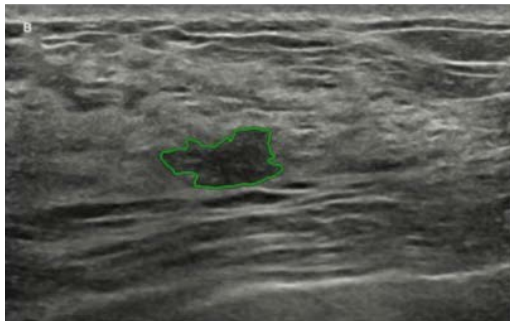
표 4. 진단 모델 개발군과 검증군에서의 BI-RADS 단독 판정과 노모그램 적용의 위양성율, 조직검사율, 민감도 성적비교

	False positive rate (%)		Biopsy Rate (%)		Sensitivity (%)	
	Development cohort	Validation cohort	Development cohort	Validation cohort	Development cohort	Validation cohort
Radiologist's BI-RADS $\geq$ 4A	68 (62, 73) [173/256]	97 (95, 99) [151/155]	72 (67, 77) [215/299]	98 (96, 100) [160/164]	98 (88, 100) [42/43]	100 (66, 100) [9/9]
Nomogram $\geq$ 114	31 (26, 37) [80/256]	45 (37, 52) [69/155]	41 (35, 46) [121/299]	48 (40, 55) [78/164]	95 (84, 99) [41/43]	100 (66, 100) [9/9]
P value	<.001	<.001	<.001	<.001	.317	NA

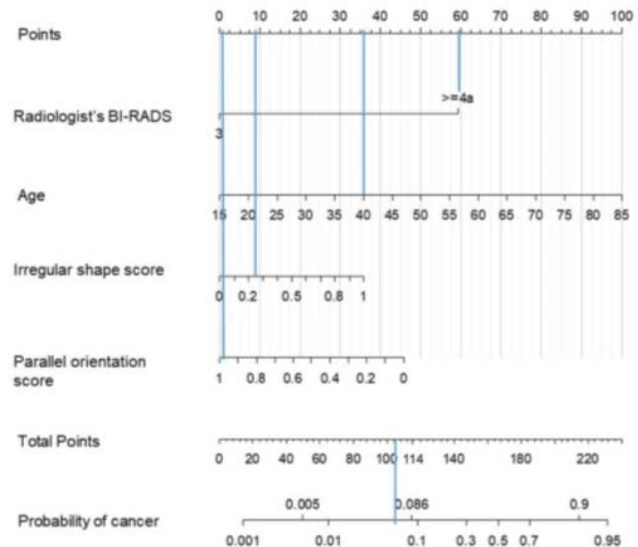
Note.- Data in parentheses are the 95% confidence intervals, and the data in brackets are the numerators/denominators.



a



b



c

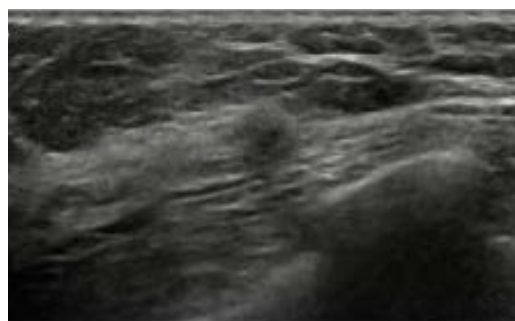
그림 5. 검진 초음파에서 발견된 40세 여성의 양성 종양 증례. (a) 1.0-cm 유방 종괴가 B-mode US에서 관찰되고, (b) 병변을 DL-CAD 소프트웨어를 이용하여 자동적으로 그려서 분석함 (초록선). 분석결과 irregular shape 점수 0.25, parallel orientation 점수 0.98로 나타났음. (c) 해당 병변의 형태학적 정량 정보에 환자 정보를 노모그램에 적용하여 점수를 산출함. 최종적으로 105점을 기록하여 기준 값 114점보다 낮은 결과로 양성에 해당하며 조직검사를 요하지 않는 것으로 평가됨.

3) 연구 결과 고찰 및 연구의 중요성

- 현재까지 DL-CAD 초음파 적용을 다룬 연구들은 검진 목적과 진단 목적의 초음파에서 발견된 종괴가 혼합되어 있었음. 따라서 검진 초음파에서 발견된 유방 종괴의 감별을 위해 DL-CAD 적용이 유용한지 평가할 수 있는 연구가 아니었음. 본 다기관 연구는 무증상이고, 유방촬영에서 이상 소견을 보이지 않았으나 유방초음파 발견된 유방 종괴가 있는 여성만을 대상으로 하여

평가하였다는 점에서 검진 환경에 가장 근접한 상황을 구현하여 그 결과를 평가할 수 있었음.

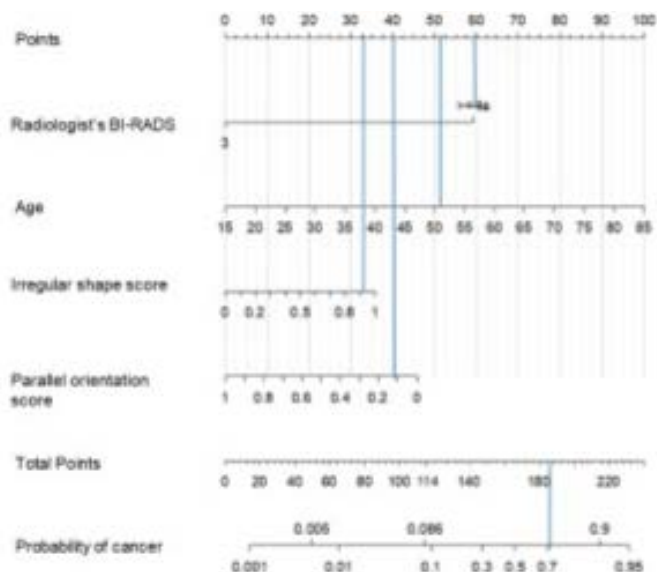
- 기존의 초음파 CAD 연구는 소프트웨어가 최종적으로 보여주는 출력 값(output) 만을 분석하였음. 화면의 표시된 출력 값은 복잡한 정량 값을 이분화하여 추려낸 데이터로, 출력과정에서 결과의 비약이 존재할 수 있음. 우리 연구에서는 이러한 중간 과정상의 결과 변형을 막기 위해 원자료 (raw data)를 이용하여 분석하였으며 검진 초음파 상황에 최적화된 연구결과를 도출하였음.



a



b



c

그림 6. 검진 초음파에서 발견된 51세 여성의 암 증례. (a) 0.9-cm 유방 종괴가 B-mode US에서 관찰되고, (b) 병변을 DL-CAD 소프트웨어를 이용하여 자동적으로 그려서 분석함 (초록선). 분석결과 irregular shape 점수 0.90, parallel orientation 점수 0.12로 나타났음. (c) 해당 병변의 형태학적 정량 정보에 환자 정보를 노모그램에 적용하여 점수를 산출함. 최종적으로 185 점을 기록하여 기준 값 114점보다 높은 결과로 악성을 시사하며 조직검사를 진행해야 하는 상황으로 평가됨.

- 기존의 초음파 DL-CAD를 다룬 연구들은 모두 단일 기관의 연구였음. 그러나 본 연구는 다기관 연구로서, 두 기관의 데이터로 최적의 진단 조합을 개발하고, 이 조합이 다른 기관의 데이터에서도 잘 적용되는지 평가하였음. 이러한 연구방식을 채택함으로써 연구 결과의 신뢰도가 높아졌음. 또한 다수의 연구자가 연구에 참여하므로 개인 연구자의 편견이 작용할 가능성이 적고, 연구 결과의 일반화에 유리함.
- 본 연구에서 초음파에서 발견된 유방암의 형태학적 정보 중 가장 중요한 특징은 병변의 모양 (irregular shape의 정도) 와 방향 (not parallel orientation) 이었음. 이러한 연구결과는 이전의 Hong 등의 연구와 Elverici 등의 연구에서 보였던 결과와 일치함. 실제로 초음파 검진에서 발견되는 병변들은 크기가 작은 병변들인데, 이러한 작은 병변은 크기가 큰 병변에 비해 변연부(margin) 평가는 제한되며, 초음파 lexicon에서 언급하는 다른 형태학적 특징을 잘 보여주지 않음.
- 본 다기관 연구를 통해서 DL-CAD 결과를 영상의학과 의사의 판단에 병용하여 사용할 경우, 검진 초음파에

서 발견된 유방 종괴의 진단성적을 높일 수 있음을 확인하였음. 특히 암을 놓치지 않으면서 위양성 조직검사를 효과적으로 줄일 수 있음을 보여줌으로써 불필요한 조직검사, 의료비 손실을 줄이는데 크게 기여할 것으로 기대됨.

- 본 딥러닝 기반 초음파 CAD 기술은 도플러와 탄성초음파가 적용되기 어려운 자동 유방 초음파 영역에도 적용 가능하며, 이 연구에 사용된 인공지능 및 딥러닝 기술은 향후 다른 장기 초음파 영상 분석 영역에 확대 적용할 수 있음.

III. 연구성과

1. 논문발표 성과정보

김수연, 최윤희, 김은경, 한부경, 윤정현, 최지수, 장정민. Deep learning-based computer-aided diagnosis in screening breast ultrasound to reduce false-positive diagnoses. Sci Rep. 2021; 11(1):395.

scientific reports



OPEN

Deep learning-based computer-aided diagnosis in screening breast ultrasound to reduce false-positive diagnoses

Soo-Yeon Kim¹, Yunhee Choi², Eun-Kyung Kim³, Boo-Kyung Han⁴, Jung Hyun Yoon³, Ji Soo Choi⁴ & Jung Min Chang^{1,5}

A major limitation of screening breast ultrasound (US) is a substantial number of false-positive biopsy. This study aimed to develop a deep learning-based computer-aided diagnosis (DL-CAD)-based diagnostic model to improve the differential diagnosis of screening US-detected breast masses and reduce false-positive diagnoses. In this multicenter retrospective study, a diagnostic model was developed based on US images combined with information obtained from the DL-CAD software for patients with breast masses detected using screening US; the data were obtained from two hospitals (development set: 299 imaging studies in 2015). Quantitative morphologic features were obtained from the DL-CAD software, and the clinical findings were collected. Multivariable logistic regression analysis was performed to establish a DL-CAD-based nomogram, and the model was externally validated using data collected from 164 imaging studies conducted between 2018 and 2019 at another hospital. Among the quantitative morphologic features extracted from DL-CAD, a higher irregular shape score ($P = .018$) and lower parallel orientation score ($P = .007$) were associated with malignancy. The nomogram incorporating the DL-CAD-based quantitative features, radiologists' Breast Imaging Reporting and Data Systems (BI-RADS) final assessment ($P = .014$), and patient age ($P < .001$) exhibited good discrimination in both the development and validation cohorts (area under the receiver operating characteristic curve, 0.89 and 0.87). Compared with the radiologists' BI-RADS final assessment, the DL-CAD-based nomogram lowered the false-positive rate (68% vs. 31%, $P < .001$ in the development cohort; 97% vs. 45% $P < .001$ in the validation cohort) without affecting the sensitivity (98% vs. 93%, $P = .317$ in the development cohort; each 100% in the validation cohort). In conclusion, the proposed model showed good performance for differentiating screening US-detected breast masses, thus demonstrating a potential to reduce unnecessary biopsies.

Mammography is an established modality for the early detection of breast cancer and reduction of breast cancer-related mortality and morbidity¹. However, as mammography alone shows a low sensitivity in women with dense breasts, various supplementary imaging modalities have been investigated for screening breast cancer². Supplementary screening breast ultrasound (US) is a widely used adjunctive tool that allows the detection of additional breast cancers in women with negative mammography². Although the number of cancers detected varies with the prevalence of cancer, it is generally modest, with an average of 3–4 cancers per 1000 US examinations^{3–7}. As most cancers detected using screening US are small, invasive and node-negative, early-stage cancer detection and management may confer a mortality benefit^{3–7}.

Despite the well-known benefits of screening breast US, a major limitation is presented by a substantial increase in false-positive findings with low specificities and low positive predictive values⁸. The Breast Imaging Reporting and Data System (BI-RADS) lexicon for breast US has contributed to the standardization of lesion characterization, description, and reporting; moreover, it has proven effective in differentiating benign and

¹Department of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea. ²Medical Research Collaborating Center, Seoul National University Hospital, Seoul, Republic of Korea. ³Department of Radiology and Research Institute of Radiological Science, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea. ⁴Department of Radiology and Center for Imaging Science, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. ⁵email: imchangjm@gmail.com

2. 학술대회 구연발표 성과정보

김수연. Screening Ultrasonography-detected Category 4A Breast Masses with a Decision-making Support Software Based on Deep Learning as an Alternative to Biopsy. RSNA 2019 Scientific presentation

3. 홍보성과

뉴스스	의료에 AI·로봇·빅데이터 결합...스마트병원 전환 '젠걸음'	2021-03-25
한국일보	여성암 1위 유방암, AI로 진단 오류 52% 줄여	2021-01-30
헬스조선	유방암 누가 걸릴까... 인공지능이 '미리' 알려준다	2021-01-29
헬스조선	인공지능으로 유방암 진단오류 획기적으로 줄인다	2021-01-27
매일경제	인공지능으로 유방암 진단오류 확 줄인다	
뉴스스	AI 초음파 검사로 유방암 진단오류 확 줄인다	
이데일리	인공지능으로 유방암 진단오류 획기적으로 줄인다	

참 고 문 헌

- Oeffinger KC, Fontham ET, Etzioni R, et al. Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA 2015; 314(15):1599-1614.
- 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):2151-2163.
- Berg WA, Zhang Z, Lehrer D, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. JAMA 2012; 307(13):1394-1404.
- 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.
- Tagliafico AS, Calabrese M, Mariscotti G, et al. Adjunct screening with tomosynthesis or ultrasound

- in women with mammography-negative dense breasts: interim report of a prospective comparative trial. J Clin Oncol 2016; 34(16):1882-1888.
- Melnikow J, Fenton JJ, Whitlock EP, et al. Supplemental screening for breast cancer in women with dense breasts: a systematic review for the US Preventive Services Task Force. Ann Intern Med 2016; 164(4):268-278.
- Rebolj M, Assi V, Brentnall A, Parmar D, Duffy SW. Addition of ultrasound to mammography in the case of dense breast tissue: systematic review and meta-analysis. Br J cancer 2018; 118(12):1559-1570.
- Lee JM, Arao RF, Sprague BL, et al. Performance of screening ultrasonography as an adjunct to screening mammography in women across the spectrum of breast cancer risk. JAMA Intern Med 2019; 179(5):658-667.
- Abdullah N, Mesurolle B, El-Khoury M, Kao E. Breast imaging reporting and data system lexicon for US: interobserver agreement for assessment of breast masses. Radiology 2009; 252(3):665-672.
- Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. Radiology 2006; 239(2):385-391.
- American College of Radiology. Breast Imaging Reporting and Data System (BIRADS). 5th ed. Reston, Va: American College of Radiology, 2013.
- Costantini M, Belli P, Lombardi R, Franceschini G, Mul? A, Bonomo L. Characterization of solid breast masses: use of the sonographic breast imaging reporting and data system lexicon. J Ultrasound Med 2006; 25(5):649-659.
- Cho N, Moon WK, Chang JM, Yi A, Koo HR, Han B-K. Sonographic characteristics of breast cancers detected by supplemental screening US: Comparison with breast cancers seen on screening mammography. Acta Radiol 2010; 51(9):969-976.
- Bae MS, Han W, Koo HR, et al. Characteristics of breast cancers detected by ultrasound screening in women with negative mammograms. Cancer science 2011; 102(10):1862-1867.
- LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521(7553):436-444.
- Choi J-H, Kang BJ, Baek JE, Lee HS, Kim SH. Application of computer-aided diagnosis in breast ultrasound interpretation: improvements in diagnostic

- performance according to reader experience. *Ultrasonography* 2018; 37(3):217.
17. Lee J, Kim S, Kang BJ, Kim SH, Park GE. Evaluation of the effect of computer aided diagnosis system on breast ultrasound for inexperienced radiologists in describing and determining breast lesions. *Med Ultrason* 2019; 21(3):239-245.
18. Di Segni M, de Soccio V, Cantisani V, et al. Automated classification of focal breast lesions according to S-detect: validation and role as a clinical and teaching tool. *J Ultrasound* 2018; 21(2):105-118.
19. Bartolotta TV, Orlando A, Cantisani V, et al. Focal breast lesion characterization according to the BI-RADS US lexicon: role of a computer-aided decision-making support. *Radiol Med* 2018; 123(7):498-506.
20. Kim K, Song MK, Kim EK, Yoon JH. Clinical application of S-Detect to breast masses on ultrasonography: a study evaluating the diagnostic performance and agreement with a dedicated breast radiologist. *Ultrasonography*. 2017; 36(1):3-9.
21. 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):749-758.
22. Sprague BL, Stout NK, Schechter C, et al. Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. *Ann Intern Med* 2015; 162(3):157-166.
23. Lee SH, Chung J, Choi HY, et al. Evaluation of screening US-detected breast masses by combined use of elastography and color Doppler US with B-mode US in women with dense breasts: a multicenter prospective study. *Radiology* 2017; 285(2):660-669.
24. Barr RG. Future of breast elastography. *Ultrasonography*. 2019; 38(2):93-105.
25. Hong AS, Rosen EL, Soo MS, Baker JA. BI-RADS for sonography: positive and negative predictive values of sonographic features. *AJR Am J Roentgenol* 2005; 184(4):1260-1265.
26. Elverici E, Bar?a AN, Akta? H, et al. Nonpalpable BI-RADS 4 breast lesions: sonographic findings and pathology correlation. *Diagn Interv Radiol* 2015; 21(3):189.
27. Chen SC, Cheung YC, Su CH, Chen MF, Hwang TL, Hsueh S. Analysis of sonographic features for the differentiation of benign and malignant breast tumors of different sizes. *Ultrasound Obstet Gynecol* 2004; 23(2):188-193.
28. Palazon-Bru A, Folgado-de la Rosa DM, Cort?s-Castell E, L?pez-Cascales MT, Gil Guill?n VF. Sample size calculation to externally validate scoring systems based on logistic regression models. *PLoS One*. 2017; 12(5):e0176726
29. Park VY, Han K, Seong YK, et al. Diagnosis of thyroid nodules: performance of a Deep Learning convolutional neural network Model vs. Radiologists. *Sci Rep* 2019; 9(1):1-9.
30. Han S, Kang H-K, Jeong J-Y, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys Med Biol* 2017; 62(19):7714.