

컴퓨터보조분석에 의한 유방촬영상
종괴와 밀도의 방사선학적 특징
추정 및 위험도 평가

연세대학교 대학원

의 과 학 과

이 재 훈

컴퓨터보조분석에 의한 유방촬영상
종괴와 밀도의 방사선학적 특징
추정 및 위험도 평가

지도교수 김 희 중

이 논문을 석사 학위논문으로 제출함

2003년 12월 일

연세대학교 대학원

의 과 학 과

이 재 훈

이재훈의 석사 학위논문을 인준함

심사위원 _____ 인

심사위원 _____ 인

심사위원 _____ 인

연세대학교 대학원

2003년 12월 일

감사의 글

청년의 푸르른 시간동안 현실적인 측면에서 자아를 돌아보고 생애의 마지막까지 다할 존립의 의미와 가치를 찾을 수 있었던 소중한 시간이었습니다. 더 넓은 세상으로 이끌어주시고 항상 세심한 배려와 깊은 관심으로 가르쳐 주신 김희중 지도교수님께 진심으로 감사드립니다. 그리고 본 논문이 나올 수 있도록 조언해 주시고 아낌없는 격려와 충고를 아끼지 않으신 김은경 교수님께 깊은 감사를 드립니다. 바쁘신 와중에도 귀중한 시간을 내주시어 부족한 논문을 꼼꼼히 살펴주신 유선국 교수님, 박병우 교수님께도 감사의 마음을 전합니다. 그동안 석사과정 동안 곁에서 힘들고 지칠 때마다 함께 해주시고 많은 도움을 주신 정해조 선생님과 실험실 동료 연구원분들, 그리고 연구진행을 위해 따뜻한 격려와 든든한 버팀목이 되어 주신 유형식 교수님, 김기황 교수님께도 감사의 말씀을 드립니다.

한결같은 믿음과 끝없는 사랑의 눈물로 기도해주신 부모님이 계셨기에 그 어느 누구보다도 더 열심히 해야한다는 마음가짐을 가질 수 있었습니다. 항상 무엇을

마무리하는 시간에 그 다음에 무엇을 할 것이냐고 묻는 말에, 출발선에서 누가 빨랐는지는 중요하지 않으며 이제 내가 가야할 길에 발을 들여 놓았을 뿐, 앞으로 부족한 나를 노력으로 채워나갈 것이라고, 그동안 한없는 은혜를 주신 많은 분들에게 약속드리겠습니다.

마지막으로 두렵고 어려울 때마다 항상 함께하시고 굳세게 하시며 의로운 오른손으로 붙들어 주신 하나님께 감사드리며 이 논문을 바칩니다.

2003. 12.

저자 씀

차 례

그림 및 표 차례	iii
국문요약	1
I. 서 론	3
II. 재료 및 방법	6
1. 유방촬영검사의 디지털 데이터베이스	6
2. 유방종괴와 밀도의 방사선학적 특징	9
3. 영상팬텀을 이용한 원형도 예비실험	13
4. 유방촬영상 종괴와 유방조직의 분리	14
5. 유방종괴와 밀도의 특성화	15
6. 광학밀도 보정 및 정량화	17
가. HOWTEK Scanner	17
나. LUMISYS Scanner	19
7. 로지스틱 회귀분석	22
가. 범주형 자료분석을 위한 로지스틱 회귀모형	23
나. 다중 로지스틱 회귀분석	24
다. 유의성 검정	25
III. 결 과	27
1. 단변량 분석	27
가. 연령에 대한 악성종양 발생률	27
나. 밀도에 대한 악성종양 발생률	30
다. 크기에 대한 악성종양 발생률	32

라. 모양에 대한 악성종양 발생률	33
마. 변연에 대한 악성종양 발생률	35
2. 다변량 분석	37
IV. 고 찰	41
V. 결 론	44
참고문헌	46
영문요약	51

그림 차례

그림 1. DDSM web page	7
그림 2. Mammographic views	8
그림 3. Anatomy of the female breast	9
그림 4. Morphologic spectrum of mammographic masses .	10
그림 5. 3D display of mammographic mass in grey level	11
그림 6. Classification of mammographic mass shapes and margins	12
그림 7. Image phantoms for the preliminary test of circularity	13
그림 8. Experimental procedures	16
그림 9. Step-wedge phantom film	17
그림 10. A plot of mean grey level versus the measured optical density of the step-wedge phantom (HOWTEK)	19
그림 11. A plot of mean grey level versus the measured optical density of the step-wedge phantom (LUMISYS)	21
그림 12. ROC curve at a confidence level of 95%	40

표 차례

표 1. Patient population	7
표 2. Experimental results of the circularities of image phantoms	14
표 3. List of optical density versus grey level (HOWTEK)	18
표 4. List of optical density versus grey level (LUMISYS)	20
표 5. Table of age versus mammary tumor	29
표 6. Testing global null hypothesis: $BETA=0$ (age)	29
표 7. Analysis of maximum likelihood estimates (age)	29
표 8. Odds ratio estimates (age)	29
표 9. Table of optical density versus mammary tumor	31
표 10. Testing global null hypothesis: $BETA=0$ (density) ..	31
표 11. Analysis of maximum likelihood estimates (density) ..	31
표 12. Odds ratio estimates (density)	31
표 13. Table of mass size versus mammary tumor	32
표 14. Testing global null hypothesis: $BETA=0$ (size)	33
표 15. Analysis of maximum likelihood estimates (size)	33

⌘ 16. Odds ratio estimates (size)	33
⌘ 17. Table of mass circularity versus mammary tumor ..	34
⌘ 18. Testing global null hypothesis: BETA=0 (shape)	34
⌘ 19. Analysis of maximum likelihood estimates (shape) ..	35
⌘ 20. Odds ratio estimates (shape)	35
⌘ 21. Table of mass margins versus mammary tumor	36
⌘ 22. Testing global null hypothesis: BETA=0 (margins) .	36
⌘ 23. Analysis of maximum likelihood estimates (margins)	36
⌘ 24. Odds ratio estimates (margins)	36
⌘ 25. Testing global null hypothesis: BETA=0	38
⌘ 26. Analysis of maximum likelihood estimates	38
⌘ 27. Odds ratio estimates	38
⌘ 28. Association of predicted probabilities and observed responses	39

국문요약

컴퓨터보조분석에 의한 유방촬영상 종괴와 밀도의 방사선학적 특징 추정 및 위험도 평가

최근 컴퓨터 기술의 발달로 종괴의 방사선학적 소견을 바탕으로 영상을 시각적으로 분석하고자 하는 연구가 많이 이루어지고 있다. 본 연구는 컴퓨터를 이용한 수학적 형태해석에 의해 종괴의 형태적 특징을 기술하고 종괴 주위의 유방밀도를 특성화하여 악성종양의 위험도를 통계적 해석을 바탕으로 객관적 및 정량적으로 분석하는데 목적이 있다.

본 연구를 위해 South Florida 대학의 디지털 데이터베이스로부터 악성종양 249 예와 양성종양 307 예를 선택하였다. 이에 대한 모든 진단은 조직학적 검사를 통해 확진되었고 고해상도 스캐너에 의해 디지털로 변환된 촬영영상은 기본적으로 내외사위촬영과 상하촬영 모두 사용되었다. 실험은 영상분할 프로그램의 도움을 받아 X선 세기정보를 이용하여 관심영역으로부터 종괴를 추출하고 컴퓨터를 이용하여 크기와 원형도를 계산하였으며 마지막으로 종괴 주위 유방밀도의 특성을 이끌어 냈다. 뿐만 아니라, 검사당시 환자의 연령과 변연의 침상 소견에 관한 자료를 포함하고 범주화하여 각각의 변수에 대해 로지스틱 모델을 이용한 단변량 분석을 실시하고 다중 로지스틱 회귀분석을 통하여 각각의 검진대상들이 복합적으로 악성종양 발생률에 어떠한 영향을 미치는지 조사하였다.

단변량 분석 결과로부터 각각의 변수가 증가할수록 악성종양에 유의한 선형 관련성이 있음을 보여주었고 다중 로지스틱 회귀분석 결

과, 95% 신뢰구간에서 높은 유의확률을 보이며 종괴의 변연, 연령, 종괴의 크기와 모양 순으로 악성종양의 발생률에 영향을 미치고 있음을 보였다. 상대적으로 종괴 주위의 유방밀도는 낮은 유의확률을 보였다.

최근 유방암 환자 수가 계속 증가하고 있으며 의료영상저장전달시스템의 도입과 디지털 유방촬영장비의 보급으로 유방전문의의 업무량이 급격히 늘어날 것으로 예상됨으로써 본 연구결과는 유방암 검진시 병변 판독의 중요한 고려 사항이 될 수 있다. 또한 유방암 검진을 위한 복합변수의 해석 및 악성 위험도 예측을 통한 컴퓨터보조진단의 가능성을 제시함으로써, 앞으로 안정적인 검진을 보조하고 환자의 불필요한 생검의 고통과 경제적인 부담을 줄여 보다 나은 의료서비스의 질적 향상을 도모할 수 있기를 기대한다.

핵심되는 말 : 유방암, 유방촬영술, 종괴, 밀도, 컴퓨터보조분석, 다중로지스틱 회귀분석

컴퓨터보조분석에 의한 유방촬영상 종괴와 밀도의 방사선학적 특징 추정 및 위험도 평가

<지도교수 김 희 중>

연세대학교 대학원 의과학과

이 재 훈

I. 서 론

한국인 여성에서 유방암은 한국중양암 등록자료¹에 의하면 2001년 여성 암 발생 등록건수 (40,191건)에서 16.1%로 1위를 차지하였고 특히 여성에서의 유방암 발생이 계속적으로 증가하였으며 추후 지속적인 증가가 예측되고 있다. 반면, 전체 여성 암 사망자 수 (21,852명) 중 유방암에 의한 사망분율은 2.1%로 유방암이 1기에 발견될 경우는 완치율 90%, 2기에는 70-80%로 조기에 발견될수록 완치율이 높은 반면, 3기로 넘어가면 20%내외로 현저하게 떨어져 치료에 어려움이 있으므로 유방암 환자의 생존율과 직결되어 있는 유방암 조기발견이 무엇보다 중요하다².

현재 유용한 진단적 방법 중에 유방촬영술 (mammography)은 효과적이고 안정적인 방법으로 가장 널리 이용되고 있으며 조기진단에 있어 매우 중요한 의미를 가지고 있다. 유방촬영술은 유방암을 잘 발견할 수 있는 검사 방법일 뿐만 아니라 무증상 여성의 선별검사와 지속적인 관리를 위해 가장 실용적인 진단방법이다^{3,4}. 하지만 유방촬영상

종괴 (mass)가 나타날 때 유방암을 의심하게 되나 이러한 소견들은 양성질환에서도 나타날 수 있다⁵. 또한 X선 촬영에 의한 유방영상은 매우 복잡하게 이루어져 있는 유선조직 (glandular tissue or parenchyma)에 의해 모호해 질 수 있고 암조직과 주위조직 간의 X선 감쇠율 (attenuation)의 차이가 적어 병변을 육안으로 감별하기 어렵기 때문에 종괴의 특징을 객관적으로 정의하고 분류하는데 문제점을 가진다^{6,7}. 후향성연구 (retrospective studies)에 의하면 위음성률도 10-15% 정도로 유방영상의 판독에는 상당한 숙련과 경험을 필요로 하게 한다⁷. 따라서 유방촬영술이 항상 효과적일 수는 없으며 의심스러운 환자에 대해 종종 생검을 통하여 결정적으로 진단하며, 여기서 조직검사는 환자에게 많은 고통과 경제적인 비용을 부담시킨다. 이런 조직검사에서 악성종양으로 판단되는 확률이 높으면 조직검사의 의미는 충분히 가지게 되는 것이지만, 조직검사의 약 65-90%가 양성종양으로 판명되고 있기 때문에 유방촬영검사에서 의심스러운 병변을 잘 선별해내는 것이 요구 된다⁸.

이와 같이 유방촬영상 판독에 있어 충분한 자격을 갖춘 가용인원이 있다 하더라도 대부분의 병변이 악성의 특징과 무관한 경우가 많고 감별에 있어 일정한 요구 수준을 유지하기가 어려움으로 유방전문의가 많은 양의 환자 영상을 판독하기가 쉽지 않다. 이러한 문제점을 보완하기 위해 여러 기술적인 방법들이 문헌에 제시되어 왔지만 종괴의 크기, 모양의 복잡성, 불분명한 변연 (margin), 종괴 경계 밖의 밀도 (density 또는 intensity)의 다양한 변화 때문에 아직까지 임상적으로 수용할만한 결과에 이르지 못하고 있다.

최근 컴퓨터 기술의 발달로 영상을 시각적으로 분석하고자 하는 연구가 많이 이루어지고 있으며 전세계적으로 컴퓨터보조진단이 방사선

과의 모든 영역에서 시도되고 있다⁹. 컴퓨터보조진단 방법은 방사선과 의사를 보조하여 업무의 효율성을 높이기 위한 보완적 도구로써 제안되어 왔으며 병변이 의심되는 부위의 이차소견으로 사용될 수 있다¹⁰. 특히 유방촬영상 종괴의 특징적 소견을 바탕으로 컴퓨터 계산에 의해 미세한 이상징후를 정확히 감별하기 위한 많은 노력이 이루어지고 있다¹¹⁻⁴. 본 연구의 목표는 컴퓨터를 이용한 수학적 형태해석에 의해 종괴의 형태적 특징을 기술하고 종괴 주위의 유방밀도를 특성화하여 악성종양에 대한 위험도를 통계적 해석을 바탕으로 객관적 및 정량적으로 평가하는데 있다.

II. 재료 및 방법

1. 유방촬영검사의 디지털 데이터베이스

Digital Database for Screening Mammography (DDSM)¹⁵는 유방 촬영 영상분석 연구를 위해 사용할 수 있도록 공유되는 자료이다 (그림 1). 이 프로젝트는 U.S. Army Medical Research and Materiel Command의 Breast Cancer Research Program으로부터 제공되고 Massachusetts General Hospital, University of South Florida, Sandia National Laboratories의 공동연구의 노력으로 이루어졌다. 또한 Washington University School of Medicine으로부터 증례가 추가되었고 공동연구 기관으로 Wake Forest University School of Medicine을 포함하고 있다. 이 데이터베이스의 목적은 감별과 진단, 교육 또는 훈련을 보조하기 위한 컴퓨터 알고리즘 개발에 있어 연구를 육성하는데 있다. 데이터베이스는 약 2,500 개의 증례를 포함하고 주로 백인 여성의 증례로 구성되었다 (표 1). 각 증례는 환자정보 (검사당시 연령), ACR (American College of Radiology)에 근거한 유방 밀도와 병적 평가, 그리고 그와 관련된 ACR 키워드 소견을 제공한다. 또한 영상정보 (스캐너, 공간 해상도 등)를 포함해 각 환자별로 유방의 네 가지 촬영영상을 제공한다. 의심되는 병변을 포함하는 영상은 위치와 유형에 대하여 픽셀 수준에서 실질적인 정보 (ground truth)를 포함하고 있다. 악성증례는 기본적인 스크리닝 검사와 최소한 한번 이상의 조직학적 검사를 통해 확진되었고 양성증례는 스크리닝 검사에서 의심스러운 소견이 발견되지만 조직학적으로, 그리고 초음파나 다른 어떤 수단으로도 악성이 아닌 것으로 감별되었다.

Volume: cancer_06 Case: A-1171-1

```

ics_version 1.0
filename A-1171-1
DATE_OF_STUDY 18 7 1984
PATIENT_AGE 73
FILM
FILM_TYPE REGULAR
DENSITY I
DATE_DIGITIZED 2 7 1998
DIGITIZER HOWTEK 43.5
SEQUENCE
LEFT_CC LINES 5491 PIXELS_PER_LINE 3166 BITS_PER_PIXEL 12 RESOLUTION 43.5 OVERLAY
LEFT_MLO LINES 5401 PIXELS_PER_LINE 3256 BITS_PER_PIXEL 12 RESOLUTION 43.5 OVERLAY
RIGHT_CC LINES 5491 PIXELS_PER_LINE 3201 BITS_PER_PIXEL 12 RESOLUTION 43.5 NON_OVERLAY
RIGHT_MLO LINES 5491 PIXELS_PER_LINE 3421 BITS_PER_PIXEL 12 RESOLUTION 43.5 NON_OVERLAY
    
```

```

FILE: A_1171_I.LEFT_MLO.OVERLAY
TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE MASS SHAPE OVAL MARGINS SPICULATED
ASSESSMENT 5
SUBTLETY 5
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 2
BOUNDARY
    CORE
    
```

```

FILE: A_1171_I.LEFT_CC.OVERLAY
TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE MASS SHAPE OVAL MARGINS SPICULATED
ASSESSMENT 5
SUBTLETY 5
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 2
BOUNDARY
    CORE
    
```

그림 1. DDSM web page. It is showing the thumbnail version of the images for case 1171. In this case, the patient age is 73 with high breast tissue density and the abnormality is a malignant spiculated mass.

표 1. Patient population

	MGH (%)	WFUSM (%)
Asian	2.06	0.2
Black	4.12	20.4
Spanish Surname	6.55	1.8
American Indian	0.00	0.1
Other	0.75	0.1
Unknown	30.34	0.3
White	56.18	77.0

All cases in DDSM are female patients.

MGH : Massachusetts General Hospital.

WFUSM : Wake Forest University School of Medicine.

본 연구를 위해 1990년에서 1998년 사이에 검사 받은 여성을 대상으로 악성종양 249 예 (환자군)와 양성종양 307 예 (대조군)로 총 556 환자의 영상을 선택하였다. 촬영영상은 기본적으로 내외사위촬영 (mediolateral oblique view : MLO)과 상하촬영 (craniocaudal view : CC) 모두 사용되었고 (그림 2), 모든 유방촬영 필름들은 MGH와 WFUSM에서 보유하고 있는 두 가지 유형의 고해상도 스캐너에 의해 디지털 영상으로 변환되었다. 본 실험에서는 8bits/pixel의 영상을 사용하였고 종괴의 특징을 정확하게 추정하기 위해 종괴를 적절한 크기로 제한하여 이러한 범위를 벗어나 미세하거나 큰 종괴는 실험대상에서 제외하였다. 그리고 종괴가 흉근 (pectoralis) 영역에 있고 MLO 또는 CC 중 어느 한 촬영상에서만 관찰되는 경우와 실질왜곡 (architectural distortion) 모양의 소견을 보이는 종괴도 제외하였다.

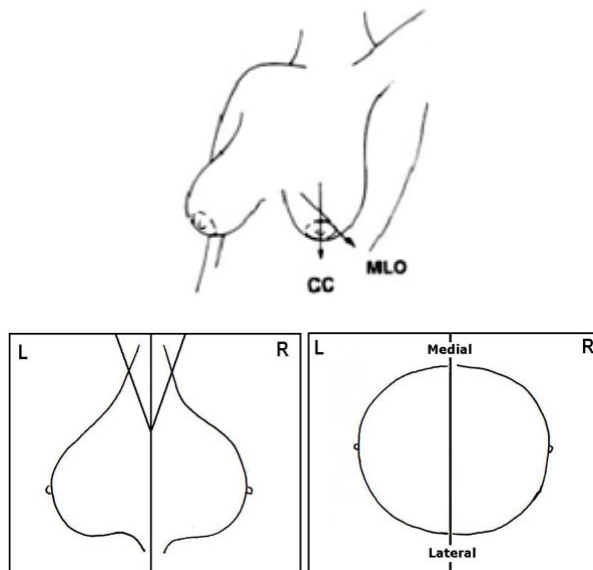


그림 2. Mammographic views. The four standard views from each case were digitized on one of two different scanners.

2. 유방종괴와 밀도의 방사선학적 특징

유방영상에서 유방조직 내에 위치하는 종괴는 위치, 모양, 밀도, 크기, 변연의 특성에 있어 매우 다양한 차이를 보이고 X선 투과량에 따라 주위조직에 비해 높은 음영으로 관찰될 수 있으며 보통 그의 진단은 방사선학적 특징을 근거로 감별될 수 있다 (그림 3)^{16,17}. 유방촬영소견상 유방암의 일차소견으로는 불규칙한 유방실질의 증가된 음영이나 종괴음영, 종괴나 비정상 증가음영 내에 한곳에 국한된 (clustered) 미세석회화침착, 종괴 주위에 짧은 (stellate), 긴 (spiculation) 침상소견이 부분적이거나 전반적으로 보이는 것이고, 이차소견으로는 유방실질의 왜곡, 유방혈관의 확장소견, 유방암 주위 지방층이 끌려가서 V자 모양을 보이는 tent sign, 유방암 주위의 Cooper's씨 인대의 비후, 전반적으로 유방음영증가소견 등을 들 수 있다¹⁸.

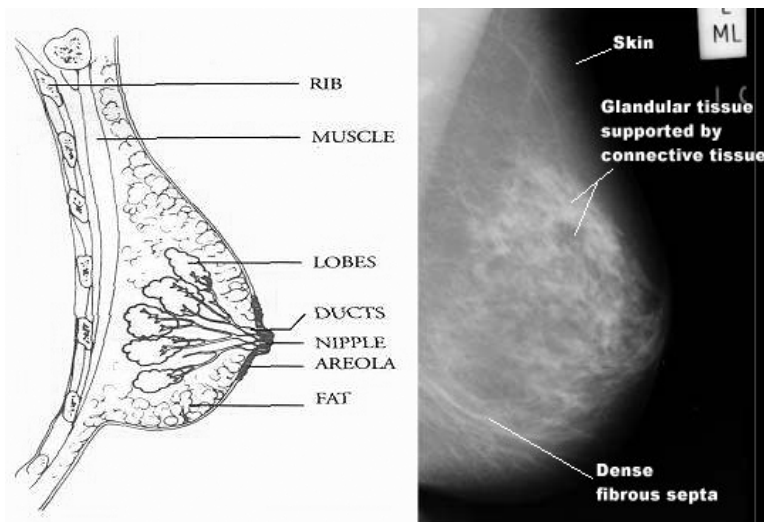


그림 3. Anatomy of the female breast. It is postulated that most cancers arise in the interlobular duct unit, either inside or just proximal to the lobule.

진단 당시 종괴의 크기는 예후를 결정하는 중요한 요소로, 전통적으로 1기와 2기를 구분 짓는 종괴 크기는 2cm로 되어 있지만 2cm 이하에서도 종괴 크기가 작을수록 예후는 좋다¹⁹. 물론 침상 변연을 가진 종괴나 경계가 좋지 않은 종괴에 대해서는 크기를 결정하기 어렵다.

대부분의 양성 유방종괴는 주위 정상조직으로부터 치밀하고 뚜렷하며 그 경계가 잘 구분되는 반면, 악성종괴는 유방암이 주변조직을 침습하고 주변조직이 암성 부위로 당겨져 왜곡을 일으키므로 보통 선명하지 않은 경계와 종괴에서 뻗어나가는 섬유성 돌출 (fibrous projection)들에 의해 불규칙한 침상 (spiculated)의 방사상 (radiating) 양상을 보인다^{20,21}. 따라서 병변의 경계가 둥근 모양에서 불규칙한 모양으로 보일수록 악성종양일 가능성이 높아진다 (그림 4)²².

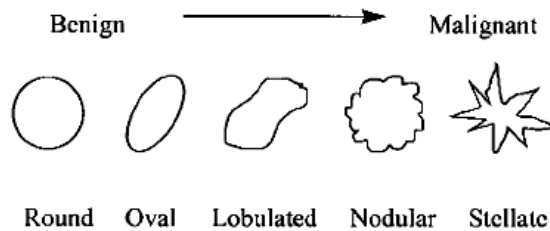


그림 4. Morphologic spectrum of mammographic masses. The shapes of mammographic masses could be used in the classification of the masses as either benign or malignant.

밀도에 있어서 대부분의 악성종괴는 유방촬영상 정상조직보다 높은 밀도를 보이며 (그림 5), 이는 암종과 연관된 강한 섬유화에 의한 것으로 생각되고 있다²³. 그러나 때로는 유방암도 등밀도로 보일 수 있고 마찬가지로 양성 병변도 높은 밀도를 보일 수 있다²⁴. 그리고 가운데가 투명한 (lucent-centered) 병변은 양성을 지시하고 비대칭성 밀도는 정상 유방조직과 구분하기 어렵다²⁵.

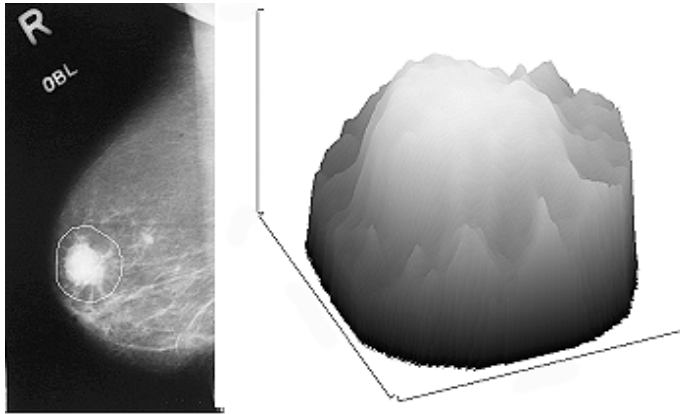


그림 5. 3D display of mammographic mass in grey level.
The high iso-intensity contours corresponds to the mass extracted from the region of interest (ROI).

유방질환의 방사선학적 특징분석은 ACR의 Breast Imaging Reporting and Data System (BI-RADS™)에 의해 다음과 같이 분류한다 (그림 6)²⁶.

가. 종괴의 모양 (shape)

- 1) 둥근형 (round)
- 2) 타원형 (oval)
- 3) 다소엽형 (lobulated)
- 4) 불규칙한 모양 (irregular)

나. 정상조직과 종괴 사이의 경계부 특성 (margins)

- 1) 명확한 경우 (circumscribed)
- 2) 숨겨진 경우 (obscured)
- 3) 미세분엽상 (microlobulated)
- 4) 불명확한 경우 또는 경계가 좋지 않은 경우 (ill-defined)
- 5) 침상 (spiculated)

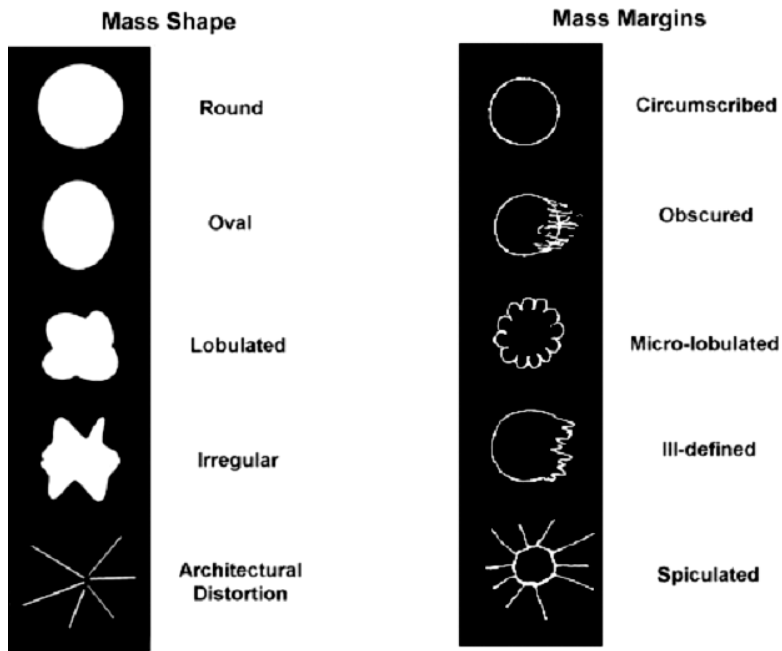


그림 6. Classification of mammographic mass shapes and margins. Architectural distortion is not technically a mass since there is no definite mass visible. Circumscribed margins are well defined and sharply demarcated with an abrupt transition between the lesion and the surrounding tissue. Obscured margins are hidden by superimposed or adjacent normal tissue. Microlobulated margins have small undulating circles along the edge of the mass. Ill-defined margins are poorly defined and scattered. Spiculated margins are marked by radiating thin lines.

다. 유방촬영상 밀도 (density or attenuation)

- 1) 고밀도 (high)
- 2) 등밀도 (equal or isodense)
- 3) 저밀도 (lower attenuation, but not fat containing)
- 4) 지방밀도 (fat containing or radiolucent)

3. 영상팬텀을 이용한 원형도 예비실험

종괴의 특징을 분석하기 위해 제안되는 컴퓨터 기술들은 종괴 영역의 형상 계수 (shape factor)와 질감 (texture) 측정에 관심이 집중되어 왔다²⁷. 종괴 영역의 윤곽을 바탕으로 하는 형상 측정은 명확한 경우 침상 종괴까지 형상의 복잡성 차이를 지시할 수 있다. 종괴 모양의 불규칙한 정도를 기술하기 위해 원형도 (circularity)의 수학적 개념을 도입하면 식 1과 같다²⁸.

$$Circularity = 4\pi (area/perimeter^2) \quad (1)$$

area : area of selection in square pixels

perimeter : length of the outside boundary of the selection

원형도 값 1은 완전한 원형을 지시하고, 원형도 값이 0에 접근할수록 그 대상의 다각형의 늘어짐이 증가함을 지시한다. 특히 원형도 값은 측정 대상의 크기와 대칭성에 상관없이 그 불규칙한 형상 계수를 정량적으로 지시하므로 종괴의 모양을 특성화하는데 유용하다. 예비실험을 위해 아래와 같이 원형에서 그 다각형이 늘어나고 복잡해지는 단계로 7가지의 영상팬텀을 만들고 (그림 7), 각각의 원형도를 측정하여 그 유용성을 검증하였다 (표 2).



그림 7. Image phantoms for the preliminary test of circularity. A mass shape may have one of seven characteristics : round, oval, nodular, lobulated I, lobulated II, irregular, and stellate.

표 2. Experimental results of the circularities of image phantoms

Series	Area	Perimeter	Circularity
1	5.307	8.624	0.897
2	4.967	8.706	0.823
3	5.162	9.260	0.757
4	4.001	8.237	0.741
5	3.167	7.810	0.653
6	3.897	10.034	0.486
7	2.960	14.457	0.178

4. 유방촬영상 종괴와 유방조직의 분리

본 연구에서 제시하는 방법은 밀도 추정에서 종괴의 외형에 구애받지 않고 어떤 ROI 형태라도 계산이 가능하고 실질적으로 종괴를 포함하는 조직의 밀도특성을 측정하기 때문에 간단하지만 만족스러운 추정이 가능하다. 실험은 크게 두 단계로 이루어져 있다. 우선 관심영역으로부터 추출된 종괴 영역은 종괴 크기와 모양의 원형도가 계산된다. 두 번째 단계에서는 종괴 주위의 나머지 정상조직 부위의 밀도특징을 이끌어 낸다. 손으로 직접 종괴의 경계를 짓는 것은 종괴의 변연의 특성이 손실될 수 있고 종괴 모양의 복잡성, 불분명한 변연, 침상의 성질 등으로 인하여 그 과정이 매우 어렵다. 하지만 유방촬영영상의 X선 세기의 정보가 종괴와 그 주위 정상조직을 분리하는데 이용될 수 있다. 보통 종괴는 유방촬영상 유방조직보다 높은 음영으로 관찰됨으로 영역을 국소화하고 종괴와 정상주위조직으로 분할하는데 동적으로 범위를 조절하는 간단한 문턱치화 (thresholding) 방법을 이용할 수 있다. 일단 실험대상의 경계를 위해 임계값이 결정되면 이

진영상에서 크기와 형태학적인 특징이 추정되었다.

본 논문에서는 공개 DICOM (digital imaging and communications in medicine) Viewer인 Image-J ver 1.30 (National Institutes of Health, Bethesda, Maryland, USA)을 이용하였다²⁹. 본 프로그램은 의학영상에서 DICOM을 비롯한 폭넓은 파일 형식을 지원하고 편집기능 및 영상분석을 위한 측정도구 등 일반적인 영상처리를 위한 많은 기능을 가지고 있다. 무엇보다도 많은 의학영상 분야의 개발자들에 의해 Plug-in 기능을 늘려갈 수 있다는 장점이 있다. 그 중에서 유방촬영 영상 유방조직으로부터 종괴를 분리하기 위해 Segmenting Assistant Plug-in을 사용하였다. 분리하고자 하는 대상에 ROI를 선정하고 그림 8에 보이는 도구를 이용하여 원하는 경계를 얻었다. 이 때에 원영상의 ROI의 수평수직 위치에서 grey level로써 최대최소 값을 자동으로 계산하고 슬라이더 바를 이동하면서 원하는 경계를 얻는다.

5. 유방종괴와 밀도의 특성화

512 × 512 ROI로부터 구분지어진 경계로부터 원형도를 이용해 종괴의 불규칙성을 정량화하고 그 지표를 가지고 형태학적으로 분류하였다. 이 방법은 기하학적 모양뿐만 아니라 복잡한 구조의 영상을 명백하게 정의하는데 신뢰할만한 결과를 보인다.

종괴의 크기는 선택된 종괴 영역의 픽셀 값을 0으로 만들고 자동으로 모든 x, y 좌표의 평균을 계산하여 중심을 구하고 그 중심으로부터 반지름의 길이가 계산되었다. 마지막으로 종괴주위 나머지 영역의 음영도를 계산하기 위해 픽셀 값 1에서 255까지의 범위에서 평균값을 계산하는 일련의 과정을 매크로 (macros) 기능으로 자동화하였다.

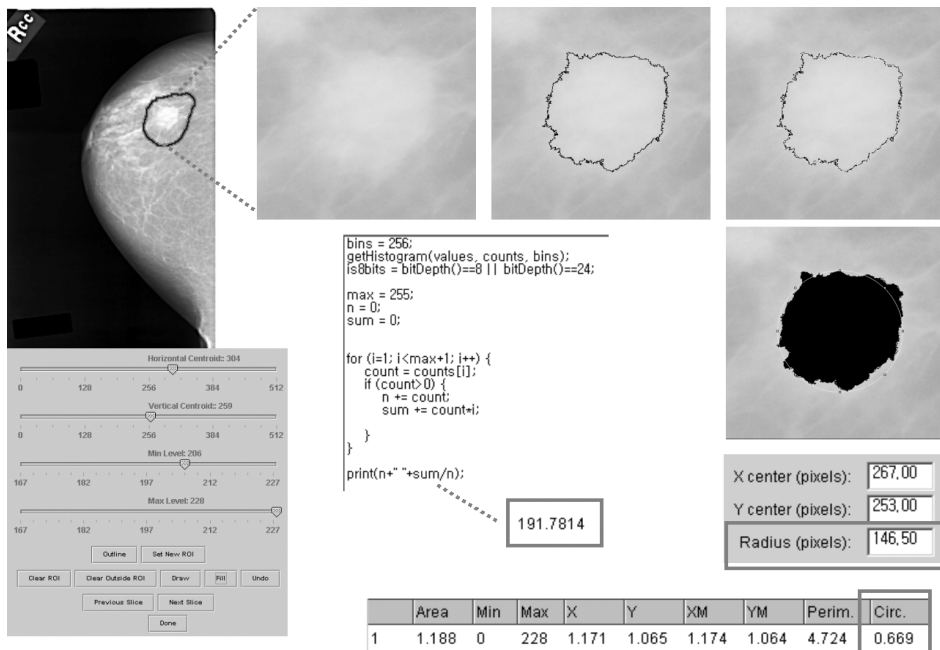
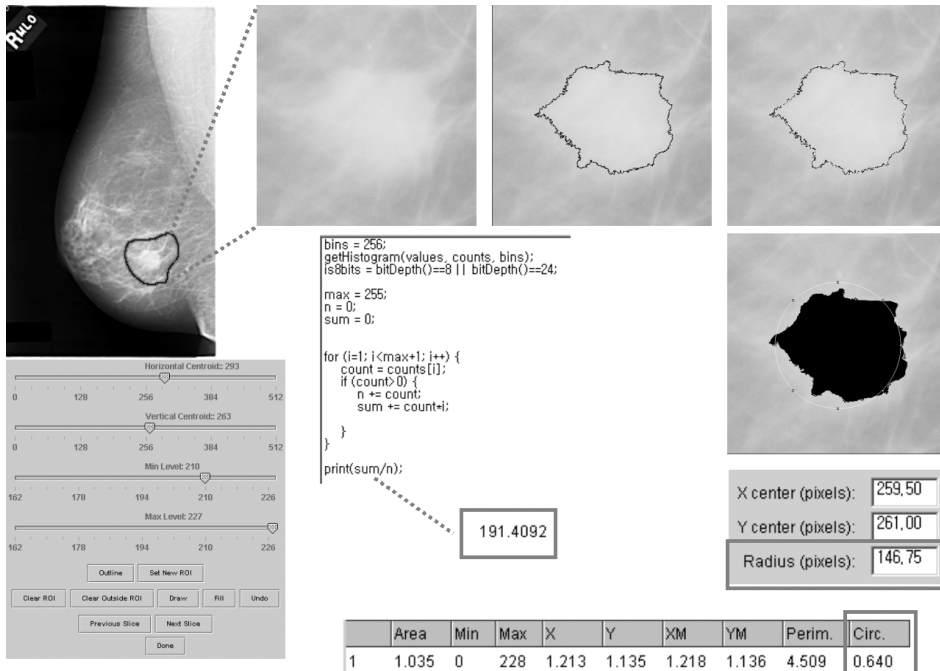


그림 8. Experimental procedures. Three radiological features were extracted from the MLO/CC images of an identical patient.

6. 광학밀도 보정 및 정량화

서로 다른 종류의 스캐너에서 얻어진 영상의 grey level은 음영도에 서로 동일한 스케일에 해당하지 않으므로 영상분석 이전에 영상 정보를 표준화 (normalize)할 필요가 있다. 종괴 주변의 정상조직의 X선 흡수 정도를 각각의 스캐너에서 step-wedge 팬텀필름으로부터 (그림 9) grey level 평균값의 선형적 변화를 보이는 광학밀도 (optical density)로 변환함으로써, 서로 다른 스캐너로부터 얻은 영상에서 표준화된 음영도를 얻어 데이터를 통합하여 처리하였다.

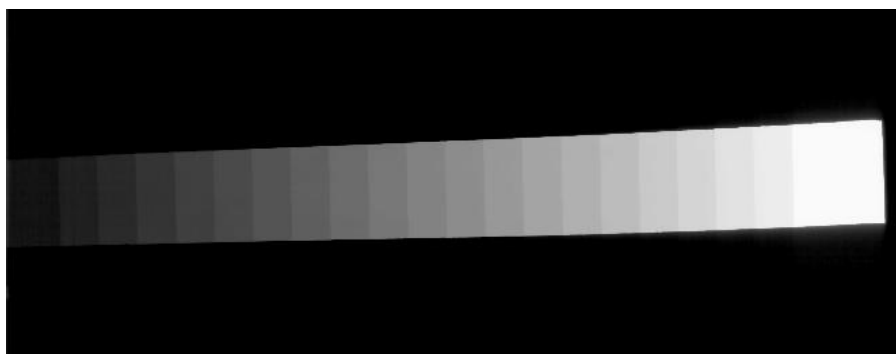


그림 9. Step-wedge phantom film. On the film, the optical density of each of the 21 wedges was measured. On the image, the grey level mean and standard deviation within each wedge was computed.

가. HOWTEK Scanner

Massachusetts General Hospital의 HOWTEK 960 스캐너에서 step-wedge 팬텀필름이 공간 해상도 픽셀당 43.5 마이크론 (microns)

으로 스캔되었다. 각 step의 광학밀도는 여러 번 측정된 후 그 평균값과 표준편차를 계산하여 표 3을 얻었다.

표 3. List of optical density versus grey level (HOWTEK)

Measured Optical Density	Grey Level Mean (Standard Deviation)
3.0600	820 (67.75)
2.9067	973 (49.15)
2.7433	1117 (33.41)
2.6033	1267 (29.57)
2.4500	1425 (24.51)
2.3033	1565 (19.21)
2.1500	1710 (18.81)
2.0000	1863 (17.52)
1.8400	2036 (13.33)
1.6900	2200 (12.65)
1.5467	2355 (10.80)
1.3967	2498 (10.69)
1.2400	2663 (11.72)
1.0933	2830 (9.38)
0.9367	2990 (9.09)
0.7900	3158 (10.83)
0.6300	3341 (9.11)
0.4800	3501 (9.37)
0.3300	3681 (10.11)
0.2100	3826 (12.25)
0.0400	4012 (4.96)

획득한 데이터에 간단한 선형회귀 곡선적합을 실시하여 절편 3.789와

기울기 -0.00094568 을 얻었다 (그림 10). 따라서 grey level 값을 광학 밀도 값으로 변환하기 위해서 step-edge 팬텀필름으로부터 얻은 광학 밀도 값 0.0400 에서 3.0600 까지의 선형적인 특성을 이용하였고 본 실험을 위해 8bits grey level 값을 12bits grey level 비율로 값을 할당 하여 광학밀도를 구할 수 있는 변환식을 만들었다 (식 2).

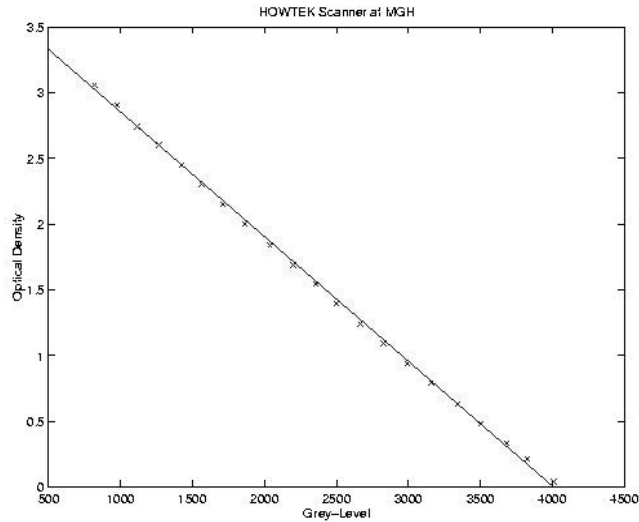


그림 10. A plot of mean grey level versus the measured optical density of the step-wedge phantom (HOWTEK).

$$\begin{aligned}
 OD_{HOWTEK} &= 3.789 - 0.00094568 \times GL_{-12bits} \\
 &= 3.789 - 0.00094568 \times (16 \times GL_{-8bits} + 8)
 \end{aligned}
 \tag{2}$$

나. LUMISYS Scanner

Wake Forest 대학의 LUMISYS 200 Laser 스캐너에서 step-wedge 팬텀필름이 공간 해상도 픽셀당 50 마이크론에서 스캔되었다. 12bits

스캐너는 0에서 4095의 픽셀 범위를 가지므로 광학밀도 0.0이 픽셀값 4095로 설정되었다. 계산된 grey level 평균값과 표준오차는 표 4와 같다.

표 4. List of optical density versus grey level (LUMISYS)

Measured Optical Density	Grey Level Mean (Standard Deviation)
0.04	4069 (14.2)
0.21	3897 (8.7)
0.34	3759 (6.2)
0.50	3601 (5.2)
0.69	3443 (5.5)
0.82	3269 (5.7)
0.96	3113 (5.2)
1.13	2948 (4.0)
1.28	2794 (6.2)
1.44	2635 (5.2)
1.59	2478 (5.2)
1.74	2324 (5.4)
1.89	2167 (5.5)
2.07	2004 (5.8)
2.21	1855 (7.6)
2.37	1690 (7.9)
2.52	1542 (9.5)
2.68	1386 (12.3)
2.84	1235 (15.4)
2.99	1099 (21.7)
3.13	957 (23.7)
3.60	495 (0.0)

step-wedge 팬텀의 grey level의 평균값에 대한 광학밀도 값을 그래프로 나타낸 것은 그림 11과 같고 계산된 데이터에 간단한 선형회귀 곡선적합을 실시하여 -1009.01의 기울기와 4096.99의 y-절편을 얻었다 (그림 11). 이로부터 선형식이 계산되었고 8bits grey level 값을 광학 밀도 값으로 변환하기 위해서 다음과 같은 식 3을 구하였다.

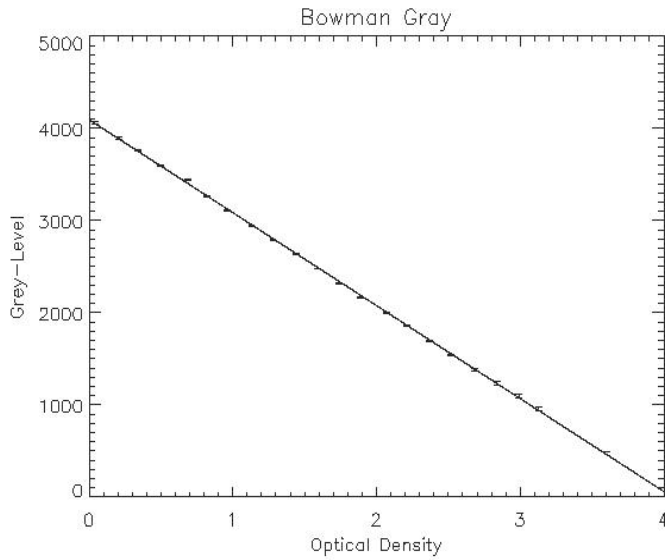


그림 11. A plot of mean grey level versus the measured optical density of the step-wedge phantom (LUMISYS).

$$\begin{aligned}
 OD_{LUMISYS} &= \frac{GL_{12bits} - 4096.99}{-1009.01} \\
 &= \frac{(16 \times GL_{8bits} + 8) - 4096.99}{-1009.01}
 \end{aligned} \tag{3}$$

7. 로지스틱 회귀분석

질병 발생 여부에 대한 자료를 회귀분석 개념 그대로 점도표로 그리면 이는 질환 발생과 각 변수의 상관성을 보여주지 못한다. 이를 해결하기 위한 방법이 독립변수에서 구간을 설정하고 각 구간에서의 질병 발생률을 구하여 이를 점도표로 표현하는 방법이다³⁰. 이렇게 변형된 곡선을 충족시켜 주는 로지스틱 함수를 이용하기 위해 다음과 같이 순서가 있는 수준으로 구별되는 범주형 자료로 분류하였다.

- 1) 광학밀도 : 1.20 미만, 1.20-1.49, 1.50-1.79, 1.80 이상의 네 범주로 분류
- 2) 종괴의 크기 : 1.00cm 미만, 1.00cm-1.49cm, 1.50cm 이상의 세 범주로 분류
- 3) 종괴의 원형도 : 0.6 미만, 0.60-0.69, 0.70-0.79, 0.80 이상의 네 범주로 분류
- 4) 환자의 연령 : 40세 미만, 40대, 50대, 60대, 70세 이상의 다섯 범주로 분류
- 5) 침상소견 : circumscribed/obscured, microlobulated/ill-defined, spiculated의 세 범주로 분류

자료에서 분류된 각 범주 (category)에 속한 개체수를 도수 (frequency)라고 하며 전체 개수에 대한 각 개체수의 비율 (proportion)로 나타낸다. 이러한 도수 혹은 비율에 대한 분석을 범주형 자료분석이라고 하는데 각각의 자료에 대하여 분할표 (contingency table) 형식으로 정리할 수 있다. 다중 로지스틱 회귀분

석을 시행하기 전에 예비검정 과정으로 로지스틱 회귀모형을 이용해 각각의 변수에 대한 분석을 독립적으로 시행하여 연구자료의 성격과 본질을 파악하였다.

가. 범주형 자료분석을 위한 로지스틱 회귀모형

이분형 질적 종속변수의 수준에서 관심의 대상이 되는 값이 나타날 확률을 P 로 표시했을 때 독립변수 x 와의 관계는 아래의 형태와 같이 로지스틱 함수 (식 4)로 표현된다³¹.

$$P(x) = \frac{1}{1 + \exp\{- (\alpha + \beta x)\}} = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \quad (4)$$

로지스틱 회귀모형에서는 종속변수의 각 수준이 일어날 확률의 오즈(odds)를 사용하여 회귀계수의 의미를 단순화시킬 수 있다. 즉 종속변수의 어떤 특정한 사건이 일어날 확률이 P 라면 일어나지 않을 확률은 $1-P$ 로 이 두 확률의 비가 오즈이고 P 와 독립변수의 관계를 앞에서 로지스틱 함수를 사용하여 설정하면 오즈는 식 5와 같이 얻어진다.

$$\begin{aligned} odds &= \frac{P}{1-P} = \frac{\exp(\alpha + \beta x)/\{1 + \exp(\alpha + \beta x)\}}{1 - [\exp(\alpha + \beta x)/\{1 + \exp(\alpha + \beta x)\}]} \quad (5) \\ &= \exp(\alpha + \beta x) \end{aligned}$$

그러나 일반 회귀모형에서는 x 의 범위가 $-\infty$ 에서 $+\infty$ 에 위치하면 y 의 범위 역시 $-\infty$ 에서 $+\infty$ 까지 위치하나, 위 모형에서는 y 의 범위가 0에서 1까지 밖에 위치하지 못하고 있다. 따라서 로지스틱 함수를 아래

식과 같이 양쪽에 log를 취해 로짓변형 (logistic transformation) 시키면 결국 다음과 같은 일반적인 회귀모형과 같은 형태를 지니는 로지스틱 회귀모형 (식 6)이 얻어진다.

$$\text{logit}(P) = \log\left(\frac{P}{1-P}\right) = \alpha + \beta x \quad (6)$$

α : $x = 0$ 일 때 사건이 일어날 $\log(\text{odds})$

β : x 가 한 단위 증가할 때 사건이 일어날 $\log(\text{odds})$ 의 증가분

로지스틱 반응함수는 시그모이드 형태를 지니며 절편 α 와 기울기 β 에 따라 형태가 변하게 된다. 결국 로지스틱 회귀모형에서 추정된 회귀계수에 antilog를 취하여 나온 값이 곧 오즈비 (odds ratio)이다. 오즈비는 두 변수간의 관련성의 정도를 나타내어 주는 지수로서 역학연구에서 사용되는 중요한 지수들 중의 하나이다. 따라서 다중 로지스틱 회귀분석 (multiple logistic regression analysis)을 실시하기 전에 각각의 변수에 대한 검정으로 참조변수가 어떤 방향인지 철저한 단변량 분석을 실시하였다.

나. 다중 로지스틱 회귀분석

몇가지 독립변수들이 함께 영향을 주고 있으며 동시에 종속변수의 각 수준마다 각각 다른 분포를 나타낼 수 있으므로 여러 가지 변수들을 포함시킨 모형이 그 자료에 대한 적절한 분석을 제공하여 준다. 이러한 다변량 분석의 목표는 모형에 포함된 각 변수의 추정치에 다른 독립변수들의 분포가 다른 것과 다른 독립변수들 간의 관련성에 대하여 통계적으로 보정 (statistically adjust)한 후에 효과를 추정하

고자 하는 것이다. 이러한 개념을 지닌 다중 로지스틱 회귀모형에 적용하면 개개의 계수추정치는 모형에 포함된 다른 모든 변수들의 영향을 조정한 후의 추정치를 나타내게 한다.

단순한 로지스틱 회귀모형을 k개의 독립변수 (X_1, X_2, \dots, X_k)를 사용하여 다음과 같은 다중 로지스틱 회귀모형 (식 7)으로 확대할 수 있다³².

$$\text{logit}(P) = \alpha + \beta_1 X_1 + \dots + \beta_k X_k, \quad \beta > 0 \quad (7)$$

수집된 자료는 SAS System 통계 패키지 ver 8.02 (SAS Institute Inc., Cary, NC, USA)를 이용하여 분석되었다. 최종적으로 광학밀도, 종피의 모양과 크기, 검사당시 환자의 연령, 침상소견을 범주형 자료로 구분하여 선형 로지스틱 모델을 이용한 다중 로지스틱 회귀분석법을 통하여 각각의 검진대상들이 악성종양 발생률에 어떠한 영향을 미치는지를 조사하였다. 특히 측정된 밀도, 종피의 원형도와 크기는 MLO와 CC 영상으로부터 계산된 값들의 평균값을 범주화하였다.

다. 유의성 검정

왈드검정 (Wald test), 우도비검정 (likelihood ratio test) 및 스코어 검정 (score test)은 일반화선형모형 (generalized linear model : GLM)에서 가설 $H_0 : \beta = 0$ (성공확률이 독립변수에 무관함)에 대한 대표적인 세 가지 유의성 검정법이다. 여기에 설정된 귀무가설은 ‘각 폭로수준에 따르는 관련도 지표로 이루어지는 직선의 기울기는 0이다’라는 의미이며 이 가설을 검정함으로써 두 변량 사이에는 직선적 관계가 있다는 결론을 유도할 수 있으며, 결과적으로는 두 변수간의

양-반응 관계에 대한 검정이 수행되게 된다. 표본크기가 매우 큰 경우 이들은 유사한 성질을 갖으며 세 가지 검정통계량의 차이점은 β 의 표본분포가 정규분포에 얼마나 잘 근사되는가에 달려있다.

가장 단순한 검정법은 최대우도추정치 (maximum likelihood estimates : MLE)의 대표본 정규성을 이용해 다음과 같은 검정통계량 (식 8)을 이용한다³³.

$$z = \frac{\beta}{ASE} \quad (8)$$

ASE : asymptotic standard error (점근표준오차)

이 통계량은 $\beta = 0$ 일 때 근사적으로 표준정규분포를 따른다. 따라서 표준정규분포를 이용하여 단측 또는 양측 검정의 p -값을 구할 수 있다. 또한 z^2 은 왈드통계량 (Wald statistic)이 되고 대표본일 때 자유도 1인 카이제곱분포가 되므로 $\beta \neq 0$ 인 양측가설의 경우 p -값은 관측값보다 큰 카이제곱 확률이 된다.

두 번째 검정법은 $H_0 : \beta = 0$ 일 때 최대로그우도를 L_0 , β 에 대한 무제한의 가정 하에서 구해진 최대로그우도를 L_1 이라 할 때, 우도비검정 통계량 $-2(L_0 - L_1)$ 은 표본크기가 큰 경우 H_0 하에서 자유도 1인 카이제곱 근사분포를 따른다. 대표본일 때는 왈드통계량의 근사가 비교적 양호하지만 표본크기가 크지 않은 경우에는 우도비검정이 카이제곱분포에 더욱 잘 근사한다고 알려져 있다³³.

우도비검정과 왈드검정에서는 최대우도추정치의 계산이 필요한 반면 최대우도추정치의 계산 절차 없이 유의성 검정을 수행할 수 있는 방법이 스코어검정이다. 이 방법은 로그우도함수에 관한 도함수 (derivatives)의 분포를 이용한 것이다³⁴.

III. 결 과

1. 단변량 분석

가. 연령에 대한 악성종양 발생률

연령별 악성종양 발생률은 40세 미만은 5.5%, 40대가 25.9%, 50대가 32.3%, 60대가 61.6%, 70대 이상은 74.1%로 연령층이 높은 군으로 갈수록 악성종양 발생률이 증가함을 보였다 (표 5).

자유도 1에서 우도비카이제곱 값은 120.655 (p -value < 0.0001)로서 $H_0 : \beta = 0$ 을 기각하게 되어 회귀모형의 기울기 β 는 0이 아니며, 따라서 연령은 악성종양에 유의한 영향을 준다고 볼 수 있다. 또한 스코어통계량 및 왈드통계량의 값도 각각 111.168 (p -value < 0.0001), 95.005 (p -value < 0.0001)로 우도비검정통계량과 유사한 결론을 나타냈다 (표 6).

또한 표 7은 회귀계수 a (intercept) 및 β 에 관한 최대우도추정치와 표준오차 (standard error : SE)를 각각 나타내므로 $z = \beta / ASE = 0.836 / 0.086 = 9.7$ 이고 왈드통계량 $z^2 = 95.0$ (p -value < 0.0001)으로 연령의 계수 $\beta = 0$ 의 귀무가설이 유의수준 5%에서 기각되었으므로 연령은 악성종양 발생률에 뚜렷한 영향을 미친다고 해석할 수 있다. 또한 $\beta > 0$ 이므로 연령대가 높아질수록 예측확률 $P(x)$ 가 커진다. 여기서 절편항의 추정치 $a = -3.023$ (SE = 0.311), 기울기 추정치 $\beta = 0.836$ (SE = 0.086)이므로 연령에 대한 악성종양 발생률의 추정된 로지스틱 회귀식은 식 6에 의해 다음과 같이 표현된다.

$$\text{logit}[P(x)] = -3.023 + 0.836x$$

기울기 추정치가 표준오차의 2배보다 크므로 유의하고 악성종양일 확률은 식 4에 의해 다음과 같이 구할 수 있다.

$$P(x) = \frac{\exp(-3.023 + 0.836x)}{1 + \exp(-3.023 + 0.836x)}$$

따라서 최소 연령대인 40세 미만 ($x = 1$)일 때 악성종양일 예측확률은 다음과 같이 얻을 수 있다.

$$P(x = 1) = \frac{\exp(-3.023 + 0.836(1))}{1 + \exp(-3.023 + 0.836(1))} = 0.1009$$

마찬가지로 최대 연령대인 70세이상일 때는 $P(x = 5) = 0.7608$ 로 예측된다.

악성 ($y = 1$)의 오즈에 대한 식으로부터 x 가 1단위 증가함에 따라 오즈는 e^β 배만큼씩 곱해져서 증가함을 알 수 있다. 따라서 연령대가 한 단위 늘어날 때마다 악성일 오즈는 $\exp(\beta) = \exp(0.836) = 2.307$ 배 즉, 230.7%만큼씩 증가한다. 이 값은 오즈비 추정치와 일치한다 (표 8).

결과적으로 환자의 연령이 늘어날수록 악성종양이 발생할 오즈가 늘어나게 되며 오즈비에 대한 월드의 95% 신뢰구간은 (1.950, 2.729)로 이 구간 내에 미지인 오즈값이 있을 것이라는 것을 95% 확신할 수 있음을 의미한다.

표 5. Table of age versus mammary tumor

		유방종양		전체	
		양성	악성		
연령	40세 미만	빈도	52	3	55
		연령의 %	94.5%	5.5%	100.0%
	40대	빈도	83	29	112
		연령의 %	74.1%	25.9%	100.0%
	50대	빈도	86	41	127
		연령의 %	67.7%	32.3%	100.0%
	60대	빈도	56	90	146
		연령의 %	38.4%	61.6%	100.0%
	70세 이상	빈도	30	86	116
		연령의 %	25.9%	74.1%	100.0%
	전체	빈도	307	249	556
		연령의 %	55.2%	44.8%	100.0%

표 6. Testing global null hypothesis: BETA=0 (age)

	자유도	값	유의확률
Likelihood Ratio	1	120.655	<0.0001
Score	1	111.168	<0.0001
Wald	1	95.005	<0.0001

표 7. Analysis of maximum likelihood estimates (age)

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr> ChiSq
Intercept	1	-3.023	0.311	94.447	<0.0001
age	1	0.836	0.086	95.005	<0.0001

표 8. Odds ratio estimates (age)

Effect	Point Estimate	95% Wald Confidence Limits
age	2.307	1.950 2.729

나. 밀도에 대한 악성종양 발생률

영상의 음영도는 광학밀도와 반비례 관계이므로 악성종양 발생률이 유방조직의 음영도가 커지는 방향으로 보이기 위해 광학밀도 값이 작아지는 순으로 분류하였다. 광학밀도 1.80 이상의 낮은 음영도가 39.2%, 광학밀도 1.50에서 1.79 사이의 중하의 음영도가 40.6%, 광학밀도 1.20에서 1.49 사이의 중상의 음영도가 51.0%, 광학밀도 1.20 미만의 높은 음영도는 54.2%로 뚜렷하지는 않지만 광학밀도 값이 작아질수록 (음영도가 증가할수록) 악성종양 발생률의 증가를 보였다 (표 9).

자유도 1에서 우도비검정통계량, 스코어통계량, 왈드통계량 모두 유의성이 p -value <0.01을 보이며 밀도의 계수 $\beta = 0$ 의 귀무가설이 기각되었다 (표 10).

밀도의 추정된 회귀계수 $a = -0.759$ (SE = 0.226)이고 $\beta = 0.237$ (SE = 0.090)이므로 광학밀도 값이 작아질수록 (음영도가 증가할수록) 악성종양일 예측확률 $P(x)$ 가 커진다 (표 11). 하지만 음영도가 가장 낮을 경우 ($x = 1$)의 악성종양일 예측확률은 다음과 같다.

$$P(x = 1) = \frac{\exp(-0.759 + 0.237(1))}{1 + \exp(-0.759 + 0.237(1))} = 0.372$$

반면 음영도가 가장 높을 경우의 악성종양일 확률 $P(x = 4) = 0.547$ 로 밀도의 변화에 따른 악성종양이 발생할 확률이 증가하지만 그 변화율이 크지 않은 것을 알 수 있다.

이러한 결과에 해당하는 오즈비는 1.268로서 밀도가 한 단위 증가하면 악성종양일 오즈는 1.268배 (126.8% 증가)가 된다 (표 12).

표 9. Table of optical density versus mammary tumor

			유방종양		전체
			양성	악성	
광학밀도	1.80 이상	빈도	76	49	125
		퍼센트	60.8%	39.2%	100.0%
	1.50-1.79	빈도	124	84	208
		퍼센트	59.6%	40.6%	100.0%
	1.20-1.49	빈도	74	77	151
		퍼센트	49.0%	51.0%	100.0%
	1.20 미만	빈도	33	39	72
		퍼센트	45.8%	54.2%	100.0%
전체	빈도	307	249	556	
	퍼센트	55.2%	44.8%	100.0%	

표 10. Testing global null hypothesis: BETA=0 (density)

	자유도	값	유의확률
Likelihood Ratio	1	7.064	0.0079
Score	1	7.045	0.0079
Wald	1	6.982	0.0082

표 11. Analysis of maximum likelihood estimates (density)

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr> ChiSq
Intercept	1	-0.759	0.226	11.299	0.0008
density	1	0.237	0.090	6.982	0.0082

표 12. Odds ratio estimates (density)

Effect	Point Estimate	95% Wald Confidence Limits
density	1.268	1.063 1.512

다. 크기에 대한 악성종양 발생률

종괴의 크기별 악성종양 발생률은 1.00cm 미만인 35.5%, 1.00cm에서 1.49cm 사이의 종괴는 45.0%, 1.50cm 이상이 51.0%였다 (표 13).

자유도 1에서 우도비검정통계량 (15.207, p -value < 0.0001), 스코어통계량 (15.109, p -value = 0.0001), 왈드통계량 (14.855, p -value = 0.0001)을 보임으로써 종괴 크기의 계수 $\beta = 0$ 의 귀무가설이 기각되었고 (표 14), 종괴 크기에 관해 추정된 회귀계수 $\beta = 0.448$ (SE = 0.116)로 종괴의 크기가 커질수록 악성종양일 예측확률 $P(x)$ 가 증가함을 알 수 있다 (표 15).

오즈비 추정치에 대한 정보로 추정된 로지스틱 회귀식에서 설명변수 종괴 크기에 대한 x 값이 $x+1$ 로 변할 때 오즈 $P / (1-P) = 1.566$ 이고 (표 16), 이 값은 $\exp(\beta) = \exp(0.448)$ 과 일치한다. 바꾸어 말하면 기울기 추정치 β 는 로그오즈비를 나타낸다. 악성종양이 발생할 오즈는 종괴 크기가 한 단위 증가함에 따라 그 이전의 약 156.6%가 증가한다.

표 13. Table of mass size versus mammary tumor

		유방종양		전체	
		양성	악성		
크기	1.00cm 미만	빈도	120	66	186
		퍼센트	64.5%	35.5%	100.0%
	1.00-1.49cm	빈도	131	107	238
		퍼센트	55.0%	45.0%	100.0%
	1.50cm 이상	빈도	56	76	132
		퍼센트	49.0%	51.0%	100.0%
전체	빈도	307	249	556	
	퍼센트	55.2%	44.8%	100.0%	

표 14. Testing global null hypothesis: BETA=0 (size)

	자유도	값	유의확률
Likelihood Ratio	1	15.207	<0.0001
Score	1	15.109	0.0001
Wald	1	14.855	0.0001

표 15. Analysis of maximum likelihood estimates (size)

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr> ChiSq
Intercept	1	-1.068	0.240	19.781	<0.0001
size	1	0.448	0.116	14.855	0.0001

표 16. Odds ratio estimates (size)

Effect	Point Estimate	95% Wald Confidence Limits
size	1.566	1.246 1.966

라. 모양에 대한 악성종양 발생률

종괴 모양의 불규칙한 정도를 지시하는 원형도별로는 0.6미만의 가장 큰 왜곡을 보이는 군이 74.4%로 가장 악성종양 발생률이 높고 원형도 0.60에서 0.69 사이가 55.8%, 0.70에서 0.79 사이는 41.9%, 그리고 가장 원형에 가까운 0.8이상이 악성종양 발생률 14.6%로 현저한 감소를 보였다 (표 17).

종괴 원형도 데이터를 이용한 로지스틱 회귀분석의 검정 결과, 우도 비검정통계량, 스코어통계량, 왈드통계량 모두 큰 카이제곱 값과 높은 유의성을 보임으로써 (p -value <0.0001), $\beta = 0$ 의 귀무가설이 기각되

었다 (표 18).

회귀계수 $\beta = 0.880$ (SE = 0.092)으로 종괴의 원형도 값이 작아질수록 (종괴의 모양이 불규칙해질수록) 악성종양일 예측확률 $P(x)$ 의 기울기가 가파르게 증가함을 알 수 있다 (표 19).

오즈비 역시 2.410으로서 종괴 모양의 불규칙한 정도를 지시하는 원형도가 한 단위 증가하면 악성종양이 발생할 오즈는 2.41배 (240% 증가)로 큰 변화율을 보였다 (표 20).

표 17. Table of mass circularity versus mammary tumor

		유방종양		전체	
		양성	악성		
원형도	0.80이상	빈도	129	22	151
		퍼센트	85.4%	14.6%	100.0%
	0.70-0.79	빈도	93	67	160
		퍼센트	58.1%	41.9%	100.0%
	0.60-0.69	빈도	53	67	120
		퍼센트	44.2%	55.8%	100.0%
	0.60미만	빈도	32	93	125
		퍼센트	25.6%	74.4%	100.0%
전체	빈도	307	249	556	
	퍼센트	55.2%	44.8%	100.0%	

표 18. Testing global null hypothesis: BETA=0 (shape)

	자유도	값	유의확률
Likelihood Ratio	1	110.079	<0.0001
Score	1	104.348	<0.0001
Wald	1	91.966	<0.0001

표 19. Analysis of maximum likelihood estimates (shape)

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr> ChiSq
Intercept	1	-2.352	0.245	92.165	<0.0001
circularity	1	0.880	0.092	91.966	<0.0001

표 20. Odds ratio estimates (shape)

Effect	Point Estimate	95% Wald Confidence Limits
circularity	2.410	2.013 2.885

마. 변연에 대한 악성종양 발생률

마지막으로 변연에 대해 circumscribed/obscured 소견을 보일 경우에 악성종양 발생률 16.4%, microlobulated/ill-defined 경우가 55.2%, 침상소견 (spiculated)이 보일 경우에 95.7%로 방사선학적으로 침상소견이 매우 중요하게 악성종양을 지시하는 것으로 조사되었다 (표 21).

따라서 기울기 추정치 $\beta = 0$ 의 귀무가설이 높은 유의성을 보이며 기각되었고 (표 22), 이 때의 회귀계수 β 는 2.121 (SE = 0.171)이다 (표 23). 추정된 회귀계수를 이용한 로지스틱 회귀식으로부터 방사선학적 소견이 명확한 변연을 지시할 때의 악성종양일 예측확률 $P(x = 1) = 0.150$ 이고 분명한 침상변연이 보일 때는 $P(x = 3) = 0.925$ 로 침상소견에 따른 악성종양 발생률이 가장 뚜렷함을 보여주었다.

추정된 오즈비는 8.335로 변연에 대한 침상소견이 뚜렷해질수록 악성종양이 발생할 오즈가 급격히 증가하였다 (표 24).

표 21. Table of mass margins versus mammary tumor

		유방종양		전체	
		양성	악성		
변연	Circumscribed	빈도	225	44	269
	/Obscured	퍼센트	83.6%	16.4%	100.0%
	Microlobulated	빈도	77	95	172
	/Ill-defined	퍼센트	44.8%	55.20%	100.0%
	Spiculated	빈도	5	110	115
		퍼센트	4.3%	95.7%	100.0%
전체		빈도	307	249	556
		퍼센트	55.2%	44.8%	100.0%

표 22. Testing global null hypothesis: BETA=0 (margins)

	자유도	값	유의확률
Likelihood Ratio	1	243.680	<0.0001
Score	1	215.813	<0.0001
Wald	1	153.203	<0.0001

표 23. Analysis of maximum likelihood estimates (margins)

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr> ChiSq
Intercept	1	-3.856	0.309	155.815	<0.0001
margins	1	2.121	0.171	153.203	<0.0001

표 24. Odds ratio estimates (margins)

Effect	Point Estimate	95% Wald Confidence Limits
margins	8.335	5.958 11.662

2. 다변량 분석

연령과 변연의 경우 p -value < 0.0001 로 매우 높은 유의성을 보이고 종피의 크기와 모양의 유의성이 각각 p -value = 0.0002, p -value = 0.0304이므로 이상의 변수 모두 악성종양 발생률에 매우 유의하게 영향을 미치고 있음을 알 수 있다 (표 26). 하지만 밀도의 경우 p -value = 0.0908로 통계학적으로 5%의 유의수준을 벗어나지만 고려하지 않을 만큼 유의하지 않다고 말할 수 없다. 따라서 최종 모형은 악성종양을 $Y = 1$ 이라 하면 다음과 같은 로지스틱 회귀식과 악성종양 발생확률을 구할 수 있다.

$$\begin{aligned} \text{logit}[P(Y=1)] &= \log\left[\frac{P(Y=1)}{1-P(Y=1)}\right] \\ &= -8.495 + 0.817X_1 + 0.22X_2 + 0.658X_3 + 0.263X_4 + 1.833X_5 \end{aligned}$$

$$P(Y=1) = \frac{\exp(-8.495 + 0.817X_1 + 0.22X_2 + 0.658X_3 + 0.263X_4 + 1.833X_5)}{1 + \exp(-8.495 + 0.817X_1 + 0.22X_2 + 0.658X_3 + 0.263X_4 + 1.833X_5)}$$

부호 자체는 그대로 해석 가능하므로 모든 변수들은 단계가 올라갈수록 악성종양 발생률은 높아지며, 각 변수의 악성에 대한 위험도를 비교하기 위해 오즈비를 해석하는 것이 도움이 된다. 즉, 연령의 오즈비가 $2.264 = \exp(0.817)$ 이며 이는 나머지 변수가 일정할 때 연령대가 한 단위 높아지면 악성종양 발생확률이 2.264배 증가한다는 의미이다. 신뢰구간은 왈드검정통계량의 분포에 근거하여 (1.814, 2.827)이다. 이와 같은 해석으로, 밀도는 1.246배, 종피의 크기는 1.931배, 종피의 모양은 1.301배, 종피의 변연은 한단계 더 높아지면 발생확률의 오즈는 $\exp(1.833) = 6.253$ 배만큼 가장 크게 변한다 (표 27). 따라서 여러 변

수를 고려한 다중 로지스틱 회귀분석으로 어느 한 변수에 대해 나머지 변수가 일정할 때 한 단계 높아질 때의 오즈비를 서로 비교한 결과, 종피의 변연 (6.253배), 연령 (2.264배), 종피의 크기 (1.931배), 종피의 모양 (1.301배), 그리고 종피 주위의 유방밑도 (1.246배) 순으로 악성종양 발생률에 영향을 미치고 있음을 보였다.

표 25. Testing global null hypothesis: BETA=0

	자유도	값	유의확률
Likelihood Ratio	5	329.963	<0.0001
Score	5	267.373	<0.0001
Wald	5	157.462	<0.0001

표 26. Analysis of maximum likelihood estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr> ChiSq
Intercept	1	-8.495	0.774	120.571	<0.0001
age	1	0.817	0.113	52.080	<0.0001
density	1	0.220	0.130	2.860	0.0908
size	1	0.658	0.178	13.700	0.0002
shape	1	0.263	0.122	4.687	0.0304
margins	1	1.833	0.201	83.590	<0.0001

표 27. Odds ratio estimates

Effect	Point Estimate	95% Wald Confidence Limits	
age	2.264	1.814	2.827
density	1.246	0.966	1.609
size	1.931	1.363	2.737
shape	1.301	1.025	1.650
margins	6.253	4.221	9.263

표 28은 추정된 회귀식으로 예측할 경우에 원래의 관측치와 일치하는 정도를 서로 다른 통계량들을 통해 보여주고 있으며 일치도 (percent concordant) 90.4%, 불일치도 (percent discordant) 9.4%, 동일한 분류를 갖는 개체의 짝 (percent tied)이 0.2%이다.

표 28. Association of predicted probabilities and observed responses

Percent Concordant	90.4%	Somers' D	0.809
Percent Discordant	9.4%	Gamma	0.811
Percent Tied	0.2%	Tau-a	0.401
Pairs	76443	c	0.905

ROC (receiver operating characteristic) 곡선은 구축한 모형의 성능을 민감도 (sensitivity)와 특이도 (specificity)에 의해 판단하고자 하는 곡선이다. 예측 모형은 입력 변수에 대해 목표 변수의 예측치를 0과 1사이의 값으로 나타낸다. 따라서 ROC 곡선은 분류 기준값의 범위인 0에서 1 사이의 모든 점에 대해서 민감도와 특이도를 계산해서 만들어 내는 곡선이다.

민감도 (sensitivity) = (실제 1, 예측 1)의 빈도 / 실제 1의 빈도

특이도 (specificity) = (실제 0, 예측 0)의 빈도 / 실제 0의 빈도

그림 12는 이에 대한 ROC 곡선으로서 구축된 모형이 어떤 성능을 갖고 있는지를 나타내고 있으며 수평축에는 1-특이도가, 수직축에는 민감도가 자리잡고 있다. ROC 곡선에서 알 수 있듯이 로지스틱 회귀 분석 결과 구축한 모형이 전체적으로 위쪽에 위치해 있으며, 이는 각

각의 분류 기준값에서의 민감도와 특이도가 높음으로써 (일치도 90.4%), 로지스틱 회귀분석 모형의 정확도가 높은 것을 의미한다.

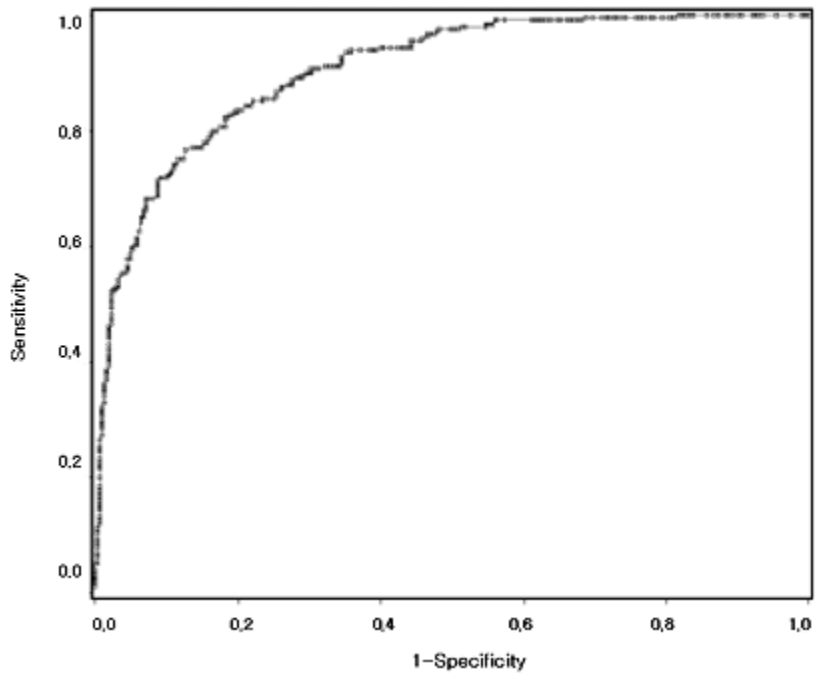


그림 12. ROC curve at a confidence level of 95%. The area under the curve (AUC), which can range from 0 to 1, is used commonly as a performance measurement. A higher AUC indicates a greater probability of making a correct decision. For this association of predicted probabilities and observed responses table, the value of the statistic c equals AUC (AUC = 0.905).

IV. 고찰

모든 변수를 포함하지 않은 각각의 변수에 대한 모형에서는 연령의 경우 오즈비 추정치는 2.307이었지만 (표 8), 나머지 변수들을 함께 포함하게 되면 2.264로 변하게 된다 (표 27). 후자의 경우는 나머지 변수들 값이 일정한 값으로 통제된 경우에 연령에 따른 추정치를 나타내므로 연령의 추정치가 다른 변수들에 관해 조정된 것으로 볼 수 있다. 이것은 단순히 연령이 악성종양 발생률에 미치는 영향만을 파악하는 것보다 다른 변수들이 반응변수에 미치는 효과를 통제하고 연령의 효과를 분석한 결과가 된다.

여러 개의 예측변수를 갖는 모형에서는 예측변수들 간에 강한 상관관계를 갖기 때문에 다중공선성 (multicollinearity)의 문제가 발생한다. 하지만 표 25, 26에서 보듯이, 임의의 어떤 변수에 대해 나머지 다른 변수들이 이미 모두 모형에 포함된 상태에서 이 변수가 모형에 추가될 필요가 있는지를 검정한 결과 높은 유의성을 보임으로써 다중공선성의 영향이 적음을 알 수 있다. 모수추정치와 표준오차를 보여주는 표 26에서 전체적인 검정은 매우 유의한 것으로 나타났지만 밀도의 유의성은 $p\text{-value} = 0.0908$ 로서 5% 유의수준을 벗어났다. 하지만 밀도의 추정치는 표준오차에 비해 매우 크므로 그 효과가 있음을 알 수 있다. 밀도에 대한 단변량 분석에서 살펴본 바에 의하면 악성종양 발생률에 대한 밀도의 효과는 $p\text{-value} = 0.0082$ 로 매우 강한 선형적 유의성을 보였지만 그 증가율이 작았던 것을 확인했었고 다중 로지스틱 회귀분석 결과 연령, 크기, 원형도, 침상소견의 변수들을 일정하게 통제할 경우 밀도의 효과는 상대적으로 매우 미미한 것으로 해석된다. 즉 여성에서 종괴가 나타날 때, 유방조직 밀도가 악성종양을 지시하

는 변수로 작용할 가능성이 적은 것으로 해석되고 단변량 분석에서 악성종양일수록 밀도의 음영도가 증가된 것은 암종과 연관된 강한 섬유화에 의한 것으로, 유방조직의 밀도와 악성종양은 직접적인 관련이 없는 것으로 생각된다. 그리고 종괴 자체의 밀도는 대체로 주위조직에 비해 높은 음영도를 보이므로 본 실험에서는 특성화하지 않았음을 밝혀둔다.

한국인 여성에서의 유방암 발생연령은 40대가 최고 호발연령이며 60세 이후가 되면 유방암의 발생빈도는 점차 감소하고 70세 이후에는 유방암의 발생 빈도가 급격하게 떨어짐으로서, 미국의 경우 50세 이후가 되면 증가하고, 70-79세 사이가 오히려 유방암의 최고 발생연령인 것과는 상당한 차이가 있다. 따라서 본 연구 대상은 대부분 백인의 미국 여성이므로 실험 결과 구축된 모형을 직접 한국 여성의 데이터에 도입할 수 없다.

현재 미국 FDA 승인 하에 몇몇 상업적 컴퓨터보조진단 시스템이 개발되어 실용화되고 있지만 엄밀히 말해 진단 (diagnosis)이 아닌 탐색발견 (detection)을 보조하는 것으로 본 실험에서 선택한 검진 대상들이 악성종양 발생률에 유의한 연관성이 있었으며 이를 토대로 통계적 과정을 통해 유방암을 예측할 수 있는 가능성을 제시한 것에 의의가 있다. 즉, 본 연구에서 제시하는 방법으로 앞으로 한국여성에 대한 유방암 데이터의 적용을 통해 컴퓨터보조진단을 위한 한 방법이 될 수 있을 것이다. 이를 위해서는 방대한 양의 유방촬영 데이터가 준비되고 활용될 수 있어야 한다는 전제가 뒤따른다. 따라서 앞으로 컴퓨터보조진단 분야의 연구 활성화를 위해 양질의 임상 데이터 관리가 필요하다.

실제로 임상에서 여러 소견들이 유방암 진단을 위해 사용되고 있

지만 본 연구에서는 유방암 진단에 가장 대표적이고 실험적으로 정량화할 수 있는 소견들을 대상으로 하였다. 따라서 컴퓨터보조진단의 정확성을 높이기 위해서 앞으로 더 많은 연구를 통해 다양한 임상소견을 추가함으로써 유방암 검진을 위한 복합변수의 해석 및 악성 위험도 예측을 통한 컴퓨터보조진단 솔루션이 완성될 수 있을 것이다.

V. 결 론

본 연구에서 사용한 통계분석 방법은 암성위험도의 예측이 가능하고 결과를 단순하고 즉시 해석이 가능하도록 비교위험도로 제시해 줄 수 있으며, 동시에 독립변수의 복합작용은 개개 위험도의 곱으로 표현할 수 있어 독립변수는 물론 복합변수가 악성종양 발생률에 미치는 영향을 해석할 수 있었다.

우리나라 생활경제수준이 향상되면서 영양과다, 만혼, 임신기피, 분만횟수 감소, 수유기피 등의 현상이 나타나고 있는 것을 감안할 때 앞으로 유방암이 더욱 증가할 가능성이 많다. 현재까지는 근본적인 원인을 찾는 역학적 연구가 많이 시행되어 일차예방도 가능하게 하기 위한 노력이 이루어지고 있지만 그보다 지속적인 유방암 선별검사로 더욱 효율적인 유방암 관리를 통해 검진에 의한 조기진단, 조기치료가 유방암 정복에 최선의 방법이 될 것이다. 즉, 적극적인 대처 방법으로 여성들이 증상이 없을 때도 규칙적으로 스크리닝 유방촬영검사 등으로 유방의 건강관리가 필요하다.

유방암의 스크리닝 검사법에서 유방촬영술은 잘 촬영하면 필름 한 장에서라도 가장 많은 소견을 찾을 수 있으며 촬영시간이 짧은 장점이 있으므로 무증상 여성의 지속적인 선별검사에 가장 적절하다. 하지만 스크리닝 검사로 탐색 중 다양한 문제로 병소를 간과하기 쉬우므로 문제점을 보완할 필요가 대두된다. 따라서 유방암의 예후를 결정하는데 발견 당시의 유방암종을 탐색발견하고 더 나아가서 상태를 명확히 기술할 필요가 있다. 예를 들어 유방암의 예후는 유방 종괴의 크기에 따라 상당한 차이가 나며 유방 종괴의 발견은 진단에 중요한 지침이 되지만 유방암 예후결정에 다양한 변수가 작용하므로 그러한

중요지침 (marker)의 연관성을 규명하는 것이 중요하다.

환자 자신들은 비정상 소견이라는 가양성 판정을 받으므로써 죽음의 공포를 갖게 되므로 판정의 정확성을 갖기 위해 기존의 이중판독 (double reading) 제도뿐만 아니라 컴퓨터보조진단에 의해 안정성과 정확성을 기하도록 하고 가능한한 판독결과를 즉시 혹은 빠른 시일 내에 환자에게 전달해 알려주어야만 주기적인 screening program에 참여하게 된다. 이것이 여성유방암 조기발견에 중요한 역할을 하리라고 생각된다. 작은 크기의 병소나 전이가 없는 조기유방암의 발견을 위해 유방촬영검사가 우수한 결과를 보이고 필수적인 역할을 하므로 컴퓨터보조진단에 의한 보완적인 유방 스크리닝 검사가 시행될 필요가 있다.

최근 유방암 환자 수가 계속 증가하고 있으며 의료영상저장전달시스템 (picture archiving and communications system : PACS)의 도입과 디지털 유방촬영장비의 보급으로 유방전문의의 업무량이 급격히 늘어날 것으로 예상됨으로써 본 연구결과는 유방암 검진시 병변 판독의 중요한 고려 사항이 될 수 있다. 또한 유방암 검진을 위한 복합변수의 해석 및 암성 위험도 예측을 통한 컴퓨터보조진단의 가능성을 제시함으로써, 앞으로 안정적인 검진을 보조하고 환자의 불필요한 생검의 고통과 경제적인 부담을 줄여서 보다 나은 의료서비스의 질적 향상을 도모할 수 있기를 기대한다.

참 고 문 헌

1. 보건복지부. 2003년 국가 암관리사업 안내; 2003.
2. Harris JR, Lippman ME, Morrow M, Hellman S. Diseases of the breast. Philadelphia: Lippincott-Raven; 1996.
3. 오기근. 유방영상학. 고려의학; 1995. p.57-84.
4. Kopans DB. Breast imaging. Philadelphia: JB Lippincott; 1989. p.1-16.
5. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the swedish two-county program of mammographic screening for breast cancer. In Bassett LW ed. RCNA : Breast imaging current status and future directions. Philadelphia: WB Saunders; 1992. p.187-210.
6. Monsees BS, Destouet JM. A screening mammography program. Staying alive and making it work. In Bassett LW ed. RCNA : Breast imaging current status and future directions. Philadelphia: WB Saunders; 1992. p.211-9.
7. Moskowitz M. Guidelines for screening for breast cancer. Is a revision in order. In Bassett LW ed. RCNA : Breast imaging current status and future directions. Philadelphia: WB Saunders; 1992. p.221-33.
8. Feig SA, Hendrick RE. Risk, benefit, and controversies in mammographic screening. In Haus AG, Yaffe MJ eds. RSNA Syllabus : A categorical course in physics. Technical aspects of breast imaging. Oak Brook: RSNA; 1994.

p.121-37.

9. Doi K, MacMahon H, Katsuragawa S, Nishikawa RM, Jiang Y. Computer-aided diagnosis in radiology: Potential and pitfalls. *Eur J Radiol* 1999;31:97-109.
10. Health M, bowyer K, Kopans D, Moore R, Kegelmeyer P Jr. Mass detection by relative image intensity. The proceedings of the 5th international workshop on digital mammography. Wisconsin: Medical Physics Publishing; 2000. p.46.
11. Doi K, Giger ML, Nishikawa RM, Hoffmann KR, MacMahon H, Schmidt RA, et al. Digital radiography: A useful clinical tool for computer-aided diagnosis by quantitative analysis of radiographic images. *Acta Radiol* 1993;34:426-39.
12. Zheng B, Chang YH, Staiger M, Good W, Gur D. Computer-aided detection of clustered microcalcifications in digitized mammograms. *Acad Radiol* 1995;2:655-62.
13. Lee CS, Kim JK, song KS, Park JM, Park HW. Computer-aided diagnosis system for the detection of microcalcifications in X-ray mammogram. *Journal of Korean Society of PACS* 1997;3:41-6.
14. Seoul National University Hospital. The first international workshop for computer aided diagnosis. 2003.
15. University of South Florida. DDSM: Digital database for screening mammography. <http://marathon.csee.usf.edu/Mammography/Database.html>.

16. Zhou C, Chan HP, Petrick N, Helvie MA, Goodsitt MM, Sahiner B, Hadjiiski LM. Computerized image analysis: Estimation of breast density on mammograms. *Med Phys* 2001;28(6):1056-69.
17. Malich A, Marx C, Facius M, boehm T, Fleck M, Kaiser WA. Tumour detection rate of a new commercially available computer-aided detection system. *Eur J Radiol* 2001;11:2454-9.
18. 연세대학교 의과대학 진단방사선과학교실. 조기유방암 진단의 접근. 1996. p.70-1.
19. RA Lorey, JL Solka, GW Rogers, DJ Marchette, CE Priebe. Mammographic computer-assisted diagnosis using computational statistics pattern recognition. *Real-Time Imaging* 1995;1:95-104.
20. Rangarai M Rangayyan, Nema M El-Faramawy, JE Leo Desautels, Onsy A Alim. Measures of acutance and shape for classification of breast tumors. *IEEE Trans on Medical Imaging* 1997;16(6):799-810.
21. Zhimin Huo, Maryellen L Giger, Carl J Vyborny, Funmi I Olopade, Dulcy E Wolverson. Computer-aided diagnosis: Analysis of mammographic parenchymal patterns and classification of masses on digitized mammograms. *Proceedings of the 20th international conference IEEE/EMBS* 1998;20(2):1017-20.
22. Lori Mann Bruce, Reza R Adhami. Classifying mammographic

- mass shapes using the wavelet transform modulus-maxima method. IEEE Trans on Medical Imaging 1999;18(12):1170-7.
23. Cynthia Y Enderwick, Evangelia Micheli-Tzanakou. Classification of mammographic tissue using shape and texture features. Proceedings of the 19th international conference IEEE/EMBS 1997:810-3.
 24. Punam K Saha, Jayaram K Udupa, Emily F Conant, Dev P Chakraborty, Daniel Sullivan. Breast tissue density quantification via digitized mammograms. IEEE Trans on Medical Imaging 2001;20(8):792-803.
 25. Lee-Han H, Cooke G Boyd NF. Quantitative evaluation of mammographic densities: A comparison of methods of assessment. Eur J Cancer Prev 1995;4(4):285-92.
 26. American College of Radiology. Breast imaging reporting and data system (BI-RADS). 3rd ed. 1998.
 27. Kim YM, Steven CH. Handbook of Medical Imaging vol. 2: Medical image processing and analysis. Washington: SPIE Press 2000. p.918-1004.
 28. John C Russ. The image processing handbook. Florida: CRC Press 1995. p.523-6.
 29. National Institutes of Health. ImageJ: Image processing and analysis in Java. <http://rsb.info.nih.gov/ij/>.
 30. 유근영. (의학-보건학을 위한) 범주형 자료 분석론. 서울대학교 출판부 1996. p.79-112.

31. 안재억, 유근영. 의학·보건학 통계분석. SPSS 아카데미 2002.
p.531-554.
32. 이성희, 황현식, 이석훈, 김정란. Win SAS V8. 교우사 2003.
p.327-336.
33. 정광모, 최용석. 범주형 자료분석 개론. 자유아카데미 2002.
p.89-130.
34. 정광모, 최용석. 로지스틱 회귀와 응용. 자유아카데미 2003.

Abstract

Radiological feature estimation and risk assessment of mammographic mass and density by computer-aided analysis

Jaehoon Lee

Department of Medical Science

The Graduate School, Yonsei University

(Directed by Professor Hee-Joung Kim)

The debate as to the best means for analyzing mammograms has now focused attention on the possible use of digital image processing and computer vision techniques to serve as an adjunct to visual analysis by a radiologist. The purpose of this study was to propose robust schemes in terms of mass diagnosis and to objectively assess the confounding effects on cancer risk.

For the experiments of mammographic masses, 249 cases of malignant (case group) and 307 cases of benign (control group) from the DDSM were selected. All diagnoses were proven by pathologic examinations and both of the two standard views, MLO and CC, were used. The produced regions of mass candidates were calculated for the irregularity and size and then densities in grey level were extracted from segmented regions for normal regions around a mass. In order to investigate the risk assessment, three

features were consequently extracted: 1) average gray value of background (converted into optical density); 2) mass size; 3) mass shape (circularity). In addition, the sets of features by the patient age at time of study and the mass margins were finally employed to statistically show the correlation with the incidence rate of malignant tumors. Each variable was categorized and then performed for the multiple logistic regression analysis to reveal the interaction to the probability of a malignant tumor.

For each variable, there were significant linear correlations with the incidence rate of malignant tumors. For the multivariate analysis with 95% confidence interval and high p -value, these analyses significantly showed that the effects on incidence rate of malignant tumors were complexly associated in the order of margins, age, mass size and shape. However density relatively showed low significance probability.

A novel approach for quantification of characteristics of mass regions in digitized mammograms was proposed and the results showed that the proposed method has a good capability to reveal the joint effect on incidence rate of malignant tumors. This method may be useful for identifying mammary tumor in mammography and developing computer-aided diagnosis as a solution.

Key Words : breast cancer, mammography, mass, density, computer-aided analysis, multiple logistic regression analysis