

딥러닝 기반 유방 x-ray 이미지 분석을 통한 유방암 분류

김동혁, 김장겸*

세종대학교, *세종대학교

miseve72@gmail.com, *jk.kim@sejong.ac.kr

Classification of breast cancer through deep learning-based breast x-ray image analysis

Kim Dong Hyeok, Kim Jang Kyum*

Sejong Univ., * Sejong Univ.

요약

본 연구는 유방 X-ray 이미지를 활용하여 유방암의 양성 및 음성 여부를 분류하는 딥러닝 기반 모델을 비교·분석하였다. 총 3,391 장의 실제 의료 데이터에 클래스 불균형과 수집 기관 간 차이를 고려한 데이터 전처리 및 증강을 적용하였고, 과적합 방지를 위해 전이학습, 드롭아웃, 학습률 스케줄링, Early Stopping 등의 전략을 도입하였다. Simulation 결과, EfficientNet-B4 가 가장 높은 정확도(Accuracy: 0.769)와 F1-score(0.696)를 기록하였으나 과적합 경향이 뚜렷하게 나타났다. 반면, YOLO v11 nano 는 안정적인 학습 곡선을 보였다.

I. 서론

유방암은 전 세계 여성에게 가장 흔하게 발생하는 암 중 하나이며, 특히 미국 여성의 전체 암 진단 사례 중 약 3 분의 1 을 차지할 정도로 높은 발병률을 보인다[1].

유방암 조기 진단을 위한 대표적인 검사 방법으로는 유방 X-ray 활영이 널리 활용되고 있다. 이 방법은 비교적 저비용이라는 장점으로 인해 보편적인 검진 수단으로 자리잡았지만, 영상 해석이 담당의의 주관적 판단, 숙련도에 따라 진단의 정확성, 일관성이 저하될 수 있는 구조적 한계를 갖는다. 이러한 문제를 보완하기 위해 컴퓨터 보조 진단(Computer-Aided Diagnosis, CAD) 시스템이 도입되었으나, 진단 성능 향상에는 한계가 있었다는 연구 결과도 보고되고 있다[2].

최근에는 인공지능 기술, 특히 딥러닝을 활용한 영상 분석 기법이 주목받고 있다. 본 연구는 유방 X-ray 이미지를 활용하여 유방암 양성, 음성을 분류하는 다양한 딥러닝 모델을 비교·분석하고, 신뢰도 높은 진단 보조 도구를 제안하고자 한다.

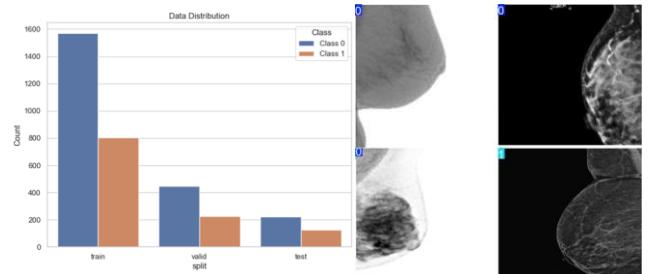
II. 본론

i. Data Characteristics and Preprocessing

본 연구에 사용된 유방 X-ray 이미지 실제 의료 환경에서 수집된 것으로, 활영 조건 및 장비의 차이로 인해 영상의 품질과 밝기에 큰 편차가 존재한다. 뿐만 아니라, 데이터 수집기관에 따른 이미지의 일관성 부족은 색 또는 좌우 반전된 이미지가 포함되는 등 다양한 모습으로 발생한다.

개인정보 보호 문제로 인해 데이터 수집이 제한적이었으며, 수집된 이미지 3391 장 중 유방암 양성 이미지 1154 장, 음성 2237 장으로 전체 클래스 비는 약 1:2 로 나타난다.

모델 학습을 위해 이를 7:2:1 비율로 분할하여 train, validation, test dataset 을 생성하였다. 이때, 각 dataset 내부 클래스 비는 1:2 로 전체 dataset 의 클래스 비와 동일하게 유지하였다.

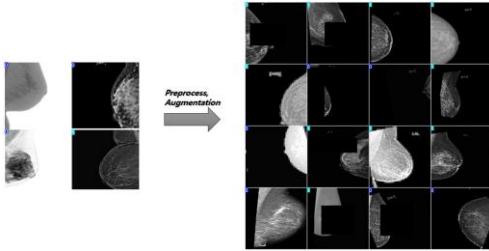


<Fig.1 left: Split and distribution, right :Sample image>

이미지 간 일관성을 확보하고자, 해상도와 밝기, 채도 등을 정규화한 후 색 반전을 일정하게 적용하여 전처리를 수행하였다.

ii. Overfitting Mitigation Strategies

해당 Task 는 데이터의 절대적인 수가 많지 않으면, 클래스 비율이 불균형하다. 이는 모델 학습에 악영향을 끼쳐 낮은 일반화 성능, 과적합(overfitting)을 발생시킬 수 있다. 이를 고려하기 위해 다양한 전략을 적용하였다. 학습 데이터의 다양성을 확보하기 위해 색상, 채도, 밝기 조절과 더불어 상하좌우 반전, 회전, 가우시안 블러 처리, 왜곡, 랜덤 Erasing 등의 기법을 적용하였다. 특히, 클래스 불균형 문제를 해소하기 위해 train date 양성 클래스 이미지는 10 배, 음성 클래스 이미지는 5 배로 증강하여 데이터셋의 균형을 맞추었다.



<Fig. 2. Image Augmentation>

모델 학습 과정에서는 ImageNet 으로 사전학습된 가중치를 모델 초기화에 활용하였다. 이러한 전이학습 기법은 적은 양의 데이터로도 빠른 수렴과 높은 일반화 성능을 기대할 수 있다는 장점이 있다.

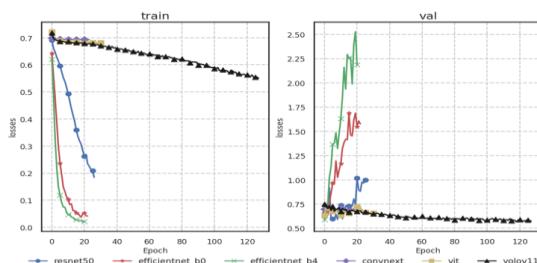
또한, 모델의 classifier 레이어에는 Dropout 기법을 0.4 의 비율로 적용하여 과적합을 억제하고자 하였다. [3]에 따르면, Dropout 은 학습 시 일부 뉴런을 임의로 비활성화함으로써 모델의 복잡도를 조절하고, 과도한 학습을 방지함으로써 일반화 성능 향상에 효과적인 방법이다.

마지막으로, Early Stopping 과 학습률 스케줄링을 적용하였다. 검증 손실이 훈련 손실보다 20 회 이상 커지면 학습이 종료시켜 불필요한 학습을 방지하였고, 학습률이 학습이 진행됨에 따라 지수적으로 감소하도록 $0.01^{(epoch/epochs)}$ 비율로 조정하였다. 이를 통해 과적합을 방지와 안정적인 수렴을 유도하고자 했다.

III. Simulation

시뮬레이션에서는 모델의 과적합 여부를 확인하기 위해 학습 곡선을 분석하였으며, 분류 성능 평가는 Accuracy, Precision, Recall, F1-score 지표를 기반으로 수행하였다. 또한, 후보군 모델로는 EfficientNet, ResNet, ConvNeXt, Vision Transformer(ViT), YOLO 를 선정했고, 모델의 구조별 성능 차이를 분석하고자 했다.

ResNet 과 EfficientNet 은 CNN 기반 모델로 이미지 분석에 많이 사용되며, Residual connection 과 Compound scaling 을 통해 성능을 개선한다. ConvNeXt 은 기존 CNN 구조에 Transformer 설계 요소를 도입해 성능을 개선했다. 반면, Vit 는 이미지를 패치로 분할한 후 self-attention 기법을 사용한다는 점에서 앞선 모델들과 차별화된다. 마지막으로, YOLO 는 객체 탐지 task 에서 주로 사용되지만, 본 연구에선 유방 X-ray 이미지를 분류하도록 구조를 수정하여 사용하였다.



<Fig. 3. Learning Curve>

Fig 3 은 각 모델의 학습 손실, 검증 손실의 변화를 시각화한 결과이다. EfficientNet-B4 는 학습 손실 값이 빠르게 감소지만, 검증 손실 값은 급격히 증가하는 양상을 보여 과적합의 경향을 보인다. 반면, YOLO v11 은 학습 손실 값의 감소는 느리지만, 검증 손실 값이 안정적으로 감소하는 모습을 보인다.

Model	Accuracy	Precision	Recall	F1-score
ResNet50	0.696049	0.601562	0.611111	0.606299
EfficientNet-B0	0.738602	0.688679	0.579365	0.62931
EfficientNet-B4	0.768997	0.701613	0.690476	0.696
ConvNeXt-tiny	0.617021	0.0	0.0	0.0
Vision Transformer-Base-16(ViT)	0.553191	0.373494	0.246032	0.296651
YOLO v11 nano	0.68997	0.633333	0.452381	0.527778

<Table 1. Models Performance>

Table 1 은 각 모델의 테스트 데이터에 대한 성능 평가를 나타낸 것이다. EfficientNet-B4 는 Accuracy 0.768997, F1-score 0.696 으로 가장 우수한 성능을 보였다. 반면, YOLO v11 는 F1-score 가 0.527778 로 상대적으로 낮았으나, 가장 낮은 검증 손실을 기록하여 학습 안정성 측면에서는 우수한 모습을 보였다. 마지막으로, ViT 와 ConvNeXt 모델은 대부분의 지표에서 낮은 수치를 보이며, 유효하게 학습되지 않은 것을 알 수 있다.

IV. 결론

본 연구에서는 유방암 X-ray 이미지를 활용하여 유방암의 양성과 음성을 분류하는 딥러닝 기반 모델들을 비교 및 분석하였다. 수집된 데이터는 클래스 불균형이 존재하고, 수집 기관 간 영상 특성에도 차이가 있어 이를 고려한 데이터 전처리 및 증강 기법과 과적합 방지 기법을 통해 모델의 일반화 성능을 높이고자 했다. 본 연구는 향후 의료기관에서 X-ray 기반 유방암 진단의 정확성을 높이고, 의료진의 진단 부담을 경감하는 데 적용될 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

이 논문은 2024년 대한민국 교육부와 한국연구재단의 인문사회분야 신진연구자지원사업의 지원을 받아 수행된 연구임(NRF-2024S1A5A802363012)

참고 문헌

- [1] C. E. DeSantis, S. A. Fedewa, A. Goding Sauer, J. L. Kramer, R. A. Smith, and A. Jemal, "Breast cancer statistics, 2015: Convergence of incidence rates between black and white women," CA: a cancer journal for clinicians, vol. 66, no. 1, pp. 31– 42, 2016.
- [2] C. D. Lehman, R. D. Wellman, D. S. Buist, K. Kerlikowske, A.N.Tosteson, D. L. Miglioretti, B. C. S. Consortium et al., "Diagnostic accuracy of digital screening mammography with and without computer aided detection," JAMA internal medicine, vol. 175, no. 11, pp. 1828– 1837, 2015.
- [3] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," The journal of machine learning research, vol. 15, no. 1, pp. 1929– 1958, 2014.