

인과 추론과 잔여 네트워크를 결합한 유방암의 림프 전이 예측

김영섭, 이하림*

금오공과대학교, *금오공과대학교

y sk980718@naver.com , *hrlee@kumoh.ac.kr

Prediction of Lymphatic Metastasis in Breast Cancer Combining Causal Inference and Residual Networks

kim Young Sub, Lee Harim*

Kumoh Univ., *Kumoh Univ.

요약

국내 유방암 환자가 지속적으로 증가함에 따라 림프절 전이 환자의 비율도 높아지고 있다. 림프절 전이는 암 전이에 중요한 원인이 된다. 그러므로 암 환자의 림프절 전이를 정확하게 예측하는 것은 중요하다. 의료진은 유방암 병리 슬라이드, 임상 결과, 영상 소견 등을 육안으로 확인하고 분석하여 유방암이 림프절로 전이 되었는지 여부를 판단하기 때문에 시간이 많이 소요된다. 따라서 본 논문에서는 유방암 림프 전이를 예측하기 위해 인과 추론과 잔여 네트워크를 결합한 딥러닝 방법을 제안한다. 임상 데이터와 병리 슬라이드를 사용하여 네트워크를 학습하고, Grad CAM을 활용하여 모델의 판단 근거를 시각적으로 제시한다. 이를 통해 의료진의 업무 효율을 높이고 판단에 도움을 주어 환자에게 빠르고 정확한 치료를 제공할 수 있도록 돕는다.

I. 서론

최근 인공지능의 성장으로 의료 분야에서도 딥러닝 기술을 기반으로 하는 지능형 질병 판단 시스템 개발이 주목을 받고 있다. 특히 유방암과 같은 질환은 조기 진단 및 전이 여부 예측이 환자 예후에 중요하다. 현재 국내 유방암 환자 수는 꾸준히 증가하고 있으며 해당 암이 림프절에 전이되는 경우 또한 증가하고 있기 때문에 해당 시스템 개발이 중요한 화두라고 할 수 있다. 림프절 전이 여부는 환자의 생존율에 큰 영향을 미치는데, 림프절 전이가 없는 경우 생존율이 98.8%이지만, 림프절 전이가 있는 경우에는 생존율이 85.5%로 감소한다. 따라서 유방암의 림프절 전이 여부는 유방암 수술 범위를 결정하고 항암 및 방사선 치료 여부를 결정하는 데 핵심적인 지표로 활용된다. 현재 림프절 전이 여부를 판단 방법은 의료진이 유방암 병리 슬라이드, 임상 결과, 영상 소견 등을 시각적으로 확인하고 판단하고 있다. 그러나 이러한 판단은 의사가 검사 결과를 분석하여 판단하기 때문에 시간이 소요되는 문제점이 존재한다. 이에 본 논문에서는 딥러닝 기술을 활용하여 유방암의 림프절 전이 여부를 판단하고 해당 판단에 있어 네트워크가 병리 슬라이드의 어떤 부분을 참고하였는지 제시하는 딥러닝 네트워크를 제안한다.

II. 유방암 환자 데이터 분석

2.1 임상 데이터 분석

임상 데이터의 항목은 결측 비율이 82% 이상인 항목들과 임상 환자의 ID 항목을 제외한 총 21개의 임상 항목들에 대해서 상관 관계도를 분석하였다. 상관 관계도를 기반으로 임상 항목을 분석하였을 때, 유방암 림프절 전이 여부 항목과 가장 관련 있는 항목은 NG(암세포와 정상 세포가 비슷한 정도), HG(암조직이 정상 조직과 비슷한 정도), HG_score_1(종양이 세관으로 구성된 정도), HG_score_2(핵의 이상 정도), DCIS_or_LCIS_여부와 T_Category(암의 병기)가 높은 상관관계를 가지는 것으로 확인하였다.

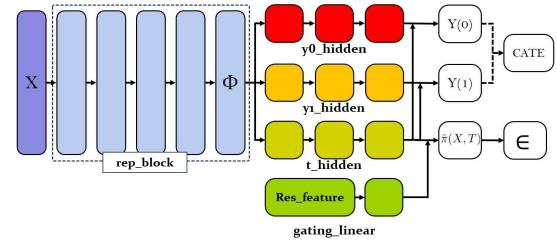


그림 1 TARNet 기반 네트워크 구조(인과 추론 모델)

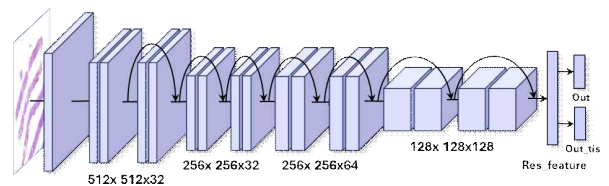


그림 2. ResNet 기반 네트워크(잔여 네트워크)

2.2 병리 슬라이드 데이터 분석

병리 슬라이드 데이터는 유방암이 림프절에 전이 된 경우에 대한 슬라이드와 유방암이 림프절에 전이되지 않은 경우에 대한 슬라이드로 이루어져 있다. 병리 슬라이드에서 나타나는 주요 병변으로는 침윤성 암, 종양, 비면포성 유관상피내암, 면포성 유관 상피 내암, 소엽성 상피 내암 등이 있다.

III. 유방암 림프절 전이 예측 모델

3.1 인과 추론 기반 임상 데이터 처리 네트워크

그림 1.은 인과 추론 기반 딥러닝 모델인 TAR 네트워크를 보여준다. 인과적 추론 모델은 기존의 딥러닝 네트워크들의 한계점이었던 원인과 결과 사이의 인과관계 파악이 부족하다는 점을 보완하기 위해 개발되었다. 특

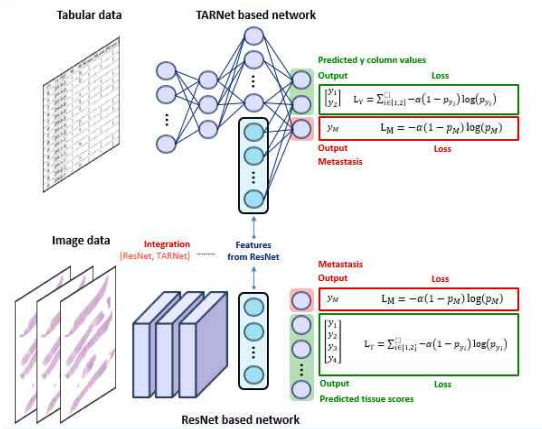


그림 3. 모델 학습 흐름도

이 이러한 인과 추론 기반 네트워크는 어떤 요인 (조치 그룹 항목)에 따라 네트워크 입력 (공변수 그룹 항목들)이 네트워크 출력 (결과 그룹 항목들)을 어떻게 변화시키는지 학습한다. 따라서 조치 그룹에 전이 여부를 두고 공변수 그룹과 결과 그룹들을 적절히 선정하면 네트워크는 전이 여부에 따른 공변수 그룹 항목들과 결과 그룹 항목들 간의 인과관계를 학습할 수 있다. 결론적으로, 해당 TAR 네트워크를 바탕으로 임상 데이터 간의 인과관계를 고려한 전이 여부 판단이 가능해진다. TAR 네트워크는 5개의 선형네트워크 층으로 구성된 Representation block과 3개의 선형네트워크 층으로 구성된 y_0 , y_1 , t_hidden 블록, gating_linear 블록으로 구성하였다.

3.2 ResNet 기반 병리 슬라이드 처리 네트워크

그림 2는 ResNet 기반 네트워크 구조이다. 이 네트워크는 이미지 데이터의 특징 추출과 패턴 인식을 위해 18개의 Convolutional Layer를 사용하여 구성하였으며, Residual Block을 사용하였다. Residual Block의 skip connection은 그림 2에서 convolution layer를 건너뛰는 화살표를 나타낸 구조로써, 입력 데이터에 대해 Convolution layer를 거치지 않고 직접 출력에 더하는 구조이다. 이 구조를 통해 역전과 과정에서 역전과 값의 감쇠를 막아 기울기 소실 문제를 해결하고 네트워크를 깊게 쌓을 수 있다.

3.3 전체 모델 및 학습 방식

그림 3은 딥러닝 모델의 훈련 과정의 전반적인 흐름을 보여준다. 전이 예측 네트워크는 정형 데이터인 표 데이터와 비정형데이터인 이미지 데이터를 동시에 사용하여 전이 여부를 판별한다. 정형데이터는 인과 추론 기반 네트워크인 TAR 네트워크에 의해서 처리되며 비정형 데이터는 Resnet 기반의 네트워크로 처리된다.

유방암의 림프절 전이 예측에 있어 정형데이터와 비정형데이터 모두를 이용하기 위해 Resnet 기반 네트워크로부터 특징 벡터를 추출하고 해당 벡터를 TAR 네트워크로 전달한다. TAR 네트워크에서는 전달받은 특징 벡터와 앞쪽 TAR 네트워크층 들로부터 추출한 특징 벡터를 혼합하고 추가 선형 네트워크 층을 통과시켜 전이 여부를 최종 판단한다. 딥러닝 모델의 훈련과정에서 TAR 네트워크는 결과 그룹과 조치 그룹의 항목을 예측하며 focal 손실 함수를 통해 네트워크가 학습된다. Resnet 기반 네트워크는 병리 슬라이드에 존재하는 병변과 림프절 전이 여부를 예측하며 이 역시 focal 손실 함수를 기반으로 학습한다.

IV. 성능 및 전이 판단 시각화

전이 예측 정확도를 유방암 림프절 전이 예측 네트워크의 성능 지표로 사

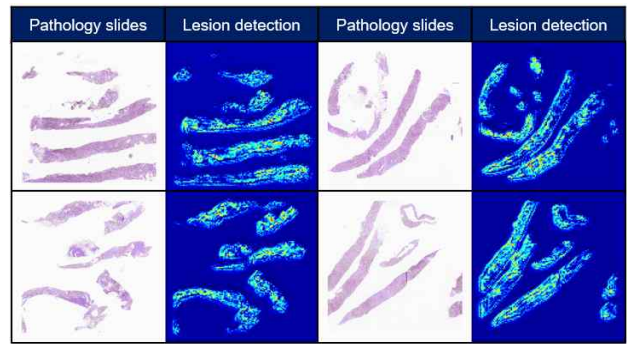


그림 4. Grad CAM을 통한 예측 근거 시각화

용하였다. 기존의 TAR 네트워크만으로 학습하였을 때는 83%의 정확도를 보였으며 TAR 네트워크에 Resnet에서 추출된 특징 벡터를 추가하여 학습하였을 때는 86%의 정확도를 보였다. 따라서 정형/비정형 데이터를 복합적으로 고려하였을 때 더 좋은 성능을 보임을 알 수 있다. 유방암이 림프절에 전이되었다고 예측한 경우, 병리 슬라이드의 특정 부분을 확인하여 예측의 근거를 시각화하는 XAI (Explainable Artificial Intelligence) 기법 중 하나인 Grad CAM (Gradient-weighted Class Activation Mapping)을 사용하였다. 그림 4의 왼쪽 이미지들은 유방암이 림프절에 전이된 병리 슬라이드를 나타내며, 오른쪽 이미지들은 왼쪽 이미지에서 어떤 부분을 고려하여 유방암 림프절 전이 여부를 예측했는지 시각적으로 표현한 것이다. 색이 붉은 정도에 따라 해당 부분이 전이 여부 예측에 큰 영향을 미쳤음을 시각적으로 확인할 수 있다.

V. 결론

본 논문에서는 유방암 림프절 전이 예측을 위해 인과적 추론과 잔류 네트워크를 결합한 방법을 제안하였다. 이를 통해 각 임상항목이 림프절 전이 여부에 미치는 영향을 해석하고, 병리 슬라이드 부터 추출된 특징을 기반으로 유방암 림프절 전이 여부를 예측하며, 예측에 대한 근거를 시각화 한다. 향후 연구 방향은 이미지 세그멘테이션 모델을 현재의 네트워크와 통합하여, 병리 슬라이드 내에서 특정 병변을 시각적으로 식별하고 표시하는 네트워크를 개발하는 것이다.

ACKNOWLEDGMENT

이 연구는 금오공과대학교 대학 연구과제비로 지원되었음(2022~2023).

참고 문헌

- [1] K. He, X. Zhang, S. Ren, J. Sun "Deep Residual Learning for Image Recognition", CoRR, Dec, 2015.
- [2] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba " Learning Deep Features for Discriminative Localization", CoRR, Dec, 2015.
- [3] B. Schölkopf, F. Locatello, S. Bauer, N. Rosemary Ke ,N. Kalchbrenner, A. Goyal, Y. Bengio "Towards Causal Representation Learning", CoRR, Feb, 2021.
- [4] Ramprasaath R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization", CoRR, Dec, 2019