

임계값 설정을 통한 근치적 위절제술 후 합병증 발생 예측 모델의 성능 평가

임수연¹, 최자윤²

¹화순전남대학교병원 간호부, ²전남대학교 간호학과

dawit21@naver.com, choijy@jnu.ac.kr

Performance of a Model to Predict Complication Occurance after Radical Gastrectomy according to Thresholds

Su-Yeon Lim¹, Ja-Yun Choi²

¹Dept. of Nursing, Chonnam National University Hwasun Hospital

²Dept. of Nursing, Chonnam National University

Abstract

위암은 전 세계적인 주요 건강문제이며, 근치적 위절제술은 위암의 표준치료이다. 근치적 위절제술 후 치료목표는 합병증 발생을 낮춰 병진 상태로 빠르게 회복하는 데 있다. 따라서, 근치적 위절제술 후 합병증 발생 여부를 선별하여 예측할 수 있는 성능이 좋은 모델을 개발하는 것은 위암환자의 회복에 매우 중요하다. 랜덤포레스트 모델은 여러 개의 결정트리를 활용한 배깅 방식의 대표적인 알고리즘으로 의료 데이터를 기반으로 한 예측에 있어 뛰어난 성능을 보여 주었다. 그러나 실제 데이터는 불균형이 빈번하게 발생하여 모델의 예측 성능에 영향을 미치므로, 최적의 분류 임계값을 설정하여 다수 클래스에 대한 편향을 줄이는 것이 중요하다. 따라서, 본 연구는 최근 10년 간 일개 대학병원의 전자의무기록 데이터를 활용하여 근치적 위절제술 후 합병증 발생을 예측하는 랜덤포레스트 모델을 개발하고, 임계값 설정을 통해 불균형 데이터에 대한 모델의 성능을 평가하고자 한다.

1. 서론

위암은 전 세계적으로 발병률과 사망률이 높은 질환으로 표준적 치료는 근치적 위절제술이다[1,2]. 근치적 위절제술은 환자의 예후와 생존률을 향상시키지만[3], 수술 후 합병증을 동반하기도 한다. 근치적 위절제술을 받은 환자의 20~24%가 수술 후 합병증을 경험하고, 그 종류와 중증도도 다양하다[4]. 이러한 합병증은 수술 후 병진 상태로의 회복을 지연시키고 환자의 예후를 악화시키며 사회·경제적 부담을 증가시켜 궁극적으로 환자의 삶의 질을 저하시킨다[5-7]. 따라서, 위암환자의 치료결과를 향상시키기 위해 수술 후 합병증 발생 여부를 선별하는 것이 중요하다[8].

머신러닝 알고리즘을 활용한 예측 모델은 의료진이 개별 환자의 위험도를 예측하고 조기관리할 수 있도록 적절한 의사결정을 지원한다[9]. 특히, 랜덤포레스트 모델은 의료 데이터를 이용한 예측에 있어 뛰어난 성능을 보고하고 있다[10, 11]. 예측 모델은 실제 데이터를 기반으로 학습할 때 복잡한 패턴과

관계를 파악하여 임상 환경에 적합한 모델로 개발된다[12]. 하지만 실제 데이터는 데이터 불균형이 흔히 발생하므로, 다수 클래스에 대한 편향을 줄이면서 기존 데이터를 수정하지 않는 임계값 설정이 중요하다. 일반적으로 분류 임계값을 0.5로 설정하는 것은 불균형 데이터에 적합하지 않으며, 모델과 데이터 특성에 맞는 최적의 임계값을 설정하는 것이 필요하다[13,14].

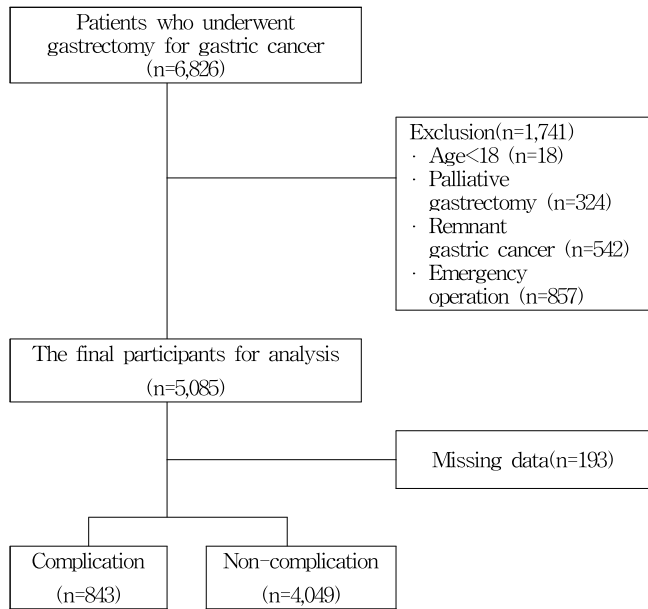
따라서 본 연구는 실제 데이터인 전자의무기록 데이터를 활용하여 근치적 위절제술 후 합병증을 예측하는 랜덤포레스트 모델을 개발하고, 적절한 임계값 설정을 통해 불균형 데이터에 대한 모델의 성능을 비교 및 평가하고자 한다.

2. 실험

2.1 데이터 세트 생성 및 전처리

본 연구 대상은 일개 대학병원에서 위암 진단 후에 근치적 위절제술을 받은 환자의 비식별화된 데이터이다. 데이터의 수집 범위는 2012년 1월부터 2022년 3월까지 10년 간의 데이터로 설정하였다. 대상자

선정을 위해 최종 진단명과 수술명을 확인하여 위절제술을 받은 위암환자 6,826건을 수집하고, 선정 및 제외기준에 따라 2차 스크리닝을 실시한 결과 총 5,085건이 수집되었다. 결과 변수인 수술 후 합병증은 퇴원 시점에 전자의무기록지에 기록된 1등급 이상의 아코디언 중증도로 정의하였다[15]. 예측 변수는 체계적 문헌고찰을 통해 수술 후 합병증 영향요인으로 확인된 변수와 전자의무기록에서 가용한 데이터를 매칭하여 41개가 선정되었다. 수집된 데이터는 결측치 193건을 제거하여 최종적으로 4,892건이 모델 개발에 사용되었으며, 수술 후 합병증 발생군은 843건(17.2%), 비발생군은 4,049건(82.8%)으로 클래스 불균형을 보였다. 구체적인 대상자 선정 과정은 그림1과 같다.

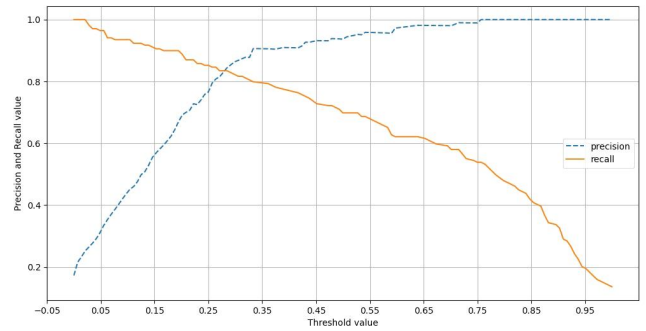


(그림1) 대상자 선정 과정

2.2 임계값 설정에 따른 성능 비교

데이터 세트는 학습용 80%, 평가용 20%의 비율로 무작위 분할하였으며, 클래스의 불균형을 고려하여 계층화된 분할을 적용하였다. 본 연구에서는 랜덤포레스트 모델의 성능 평가를 위해 정밀도(Precision), 재현율(Recall), F1 score, 정확도(Accuracy), 특이도(Specificity), ROC-AUC (Receiver Operating Characteristic Area Under the Curve) 6가지 지표를 사용하였다.

모델의 임계값 변경에 따른 정밀도-재현율 변화 곡선은 그림2와 같으며, 최적의 임계값은 0.33으로 확인되었다. 임계값 설정에 따른 모델의 성능은 최적의 임계값 0.33을 중심으로 0.13부터 0.43까지의



(그림2) 임계값 설정에 따른 정밀도-재현율 곡선 변화

범위를 0.1 간격으로 변화시키면서 성능 지표를 확인하였다. 실험 결과, 표 1과 같이 임계값이 높아질수록 양성 예측에 대한 기준이 높아지면서, 재현율이 낮아짐을 확인할 수 있었다. 임계값 설정은 ROC-AUC 값에는 영향을 미치지 않았다.

<표1> 임계값 설정에 따른 성능 비교 결과

Metrics	Threshold			
	0.13	0.23	0.33 (Best)	0.43
Precision	0.51	0.74	0.90	0.93
Recall	0.92	0.86	0.80	0.75
F1 score	0.65	0.79	0.85	0.83
Accuracy	0.83	0.92	0.95	0.95
Specificity	0.81	0.94	0.98	0.99
ROC-AUC	0.96	0.96	0.96	0.96

3. 결론

본 연구에서는 다양한 임계값 설정에 따른 근처적 위절제술 후 합병증 예측 모델의 성능을 비교 및 평가하였다. 그 결과, 최적의 임계값 설정이 환자 의무기록지와 같이 불균형한 데이터를 활용한 예측모델에서 성능을 개선할 수 있음을 확인하였다. 또한, 불균형 데이터에서 모델의 성능을 평가하는데 있어 재현율, 정밀도, F1 score가 적합한 지표임을 확인하였다.

참고문헌

[1] Sung, Hyuna, et al. "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA: a cancer journal for clinicians, 71, 3, 209-249, 2021.

[2] Orditura, Michele, et al. "Treatment of gastric cancer." World journal of gastroenterology:

- WJG 20, 7, 1635, 2014.
- [3] Panda, Sangram K., et al. "Evolution of treatment in gastric cancer—a systematic review." *Journal of the Egyptian National Cancer Institute* 34, 1, 12, 2022.
- [4] Papenfuss, Wesley A., et al. "Morbidity and mortality associated with gastrectomy for gastric cancer." *Annals of surgical oncology* 21, 3008–3014, 2014.
- [5] Wang, Shiqi, et al. "Postoperative complications and prognosis after radical gastrectomy for gastric cancer: a systematic review and meta-analysis of observational studies." *World journal of surgical oncology* 17, 1–10, 2019.
- [6] Selby, Luke V., et al. "Association of hospital costs with complications following total gastrectomy for gastric adenocarcinoma." *JAMA surgery* 152, 10, 953–958, 2017.
- [7] Karanicolas, Paul J., et al. "Quality of life after gastrectomy for adenocarcinoma: a prospective cohort study." *Annals of surgery* 257, 6, 1039–1046, 2013.
- [8] Kanda, Mitsuro. "Preoperative predictors of postoperative complications after gastric cancer resection." *Surgery today* 50, 1, 3–11, 2020.
- [9] 한형진. "의료/헬스케어 분야에서의 설명 가능 인공지능 (Explainable AI) 연구 동향." BRIC View 동향리포트, BRIC View T13, 2021.
- [10] Stam, Wessel T., et al. "The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: a systematic review." *Surgery* 171, 4, 1014–1021, 2022.
- [11] Bektaş, Mustafa, et al. "Machine learning applications in upper gastrointestinal cancer surgery: a systematic review." *Surgical endoscopy* 37, 1, 75–89, 2023.
- [12] Chandran, Urmila, et al. "Machine learning and real-world data to predict lung cancer risk in routine care." *Cancer Epidemiology, Biomarkers & Prevention* 32, 3, 337–343, 2023.
- [13] Zou, Quan, et al. "Finding the best classification threshold in imbalanced classification." *Big Data Research* 5, 2–8, 2016.
- [14] Leevy, Joffrey L., et al. "Threshold optimization and random undersampling for imbalanced credit card data." *Journal of Big Data* 10, 1, 58, 2023.
- [15] Jung, Mi Ran, et al. "Definition and classification of complications of gastrectomy for gastric cancer based on the accordion severity grading system." *World journal of surgery* 36, 10, 2400–2411, 2012.
- [16] Hassanat, Ahmad B., et al. "Stop oversampling for class imbalance learning: A critical review." *arXiv preprint arXiv:2202.03579*, 2022.