

백혈구 수치 예측을 위한 인공지능 모델에 관한 연구

한주혁, 오현우, 김웅식, 김용석*

건양대학교, *건양대학교

21856503@konyang.ac.kr, osj0805@naver.com, wskim@konyang.ac.kr, *yskim@konyang.ac.kr

A Study on the Artificial Intelligence Model for the Prediction of White Blood Cell Count

Han Ju-Hyuck, Oh Hyun-Woo, Kim Woong-Sik, Kim Yong-Suk*

Konyang Univ., *Konyang Univ.

요약

본 논문에서는 딥러닝을 이용하여, 기존의 혈액검사가 필요했던 백혈구 수치를 예측하는 인공지능 모델을 연구했다. 백혈구 수치는 체내 염증 수치 및 패혈성 쇼크를 알 수 있는 중요한 인체정보이다. 그러나 백혈구 수치는 혈액검사가 필요하며, 검사결과를 확인하기 위해서는 2시간의 의료대기 시간이 필요하다. 따라서 응급한 환자들은 백혈구 수치를 확실하게 확인할 수 있는 혈액검사 결과를 기다리는 동안에 충분한 의료적 대응을 받기 어려울 수 있다. 이러한 문제를 해결하기 위한 과정으로 의료 현장에서는 다른 생체신호 및 정보를 바탕으로 의료적 대응을 수행하지만, 정량적으로 제공되는 백혈구 수치에 따른 대응과는 차이가 있다. 따라서 본 연구에서는 인공지능의 GAN(Generative Adversarial Networks)을 기반으로 환자의 생체신호 및 정보를 입력받아, 백혈구 수치를 예측하는 인공지능 모델에 관한 연구를 수행했다. 본 연구의 인공지능 모델은 결측되지 않은 데이터를 학습하여, 백혈구 수치가 결측된 데이터가 입력으로 작용했을 때, 이를 정량적으로 예측한다. 모델의 검증은 원본 데이터와 결측 데이터를 혼합하여, 인공지능 모델에 입력했을 때의 성능으로 구성했다. 검증 결과로는 백혈구 예측모델을 통해 생성된 백혈구 수치를 포함한 혼합 데이터군이 원본 데이터군보다 패혈증 환자에 대한 검출 성능이 뛰어난 것으로 나타났다.

I. 서론

본 논문에서는 혈액검사가 필요한 백혈구 수치(White Blood Cell Count: WBC)를 인공지능의 GAN(Generative Adversarial Networks)을 기반으로 예측하는 인공지능 모델에 관해 연구했다. 백혈구 수치는 체내 염증 수치 및 패혈성 쇼크 등의 환자의 생명을 위협하는 심각한 상태를 보여주는 생체정보이다.[1] 이는 현대 의학에서 환자의 상태를 파악하기 위한 필수적인 생체정보이며, 즉각적인 의료적 대응이 필요한 응급실 및 중환자실에서 더욱 중요하다.[1][2] 그러나 백혈구 수치는 혈액검사가 필요하다. 혈액검사는 혈액에서 의학적 정보를 얻기 위한 검사로 평균 두 시간 정도의 시간 소요가 발생한다.[3] 그러므로 즉각적인 의료 대응이 필요한 응급실 및 중환자실의 환자들은 적절한 치료에 시간적 여유가 필요하다. 이러한 문제를 해결하기 위하여, 병원에서는 환자의 생체신호 및 정보를 바탕으로 혈액검사의 결과가 도출되기 전에 환자에게 의료적 대응을 수행한다. 그러나 이는 의료진의 경험적인 학습을 기반으로 하여, 상대적으로 경험이 적은 의료진은 이러한 대응에 어렵다.[4] 또한, 정량적인 혈액검사 결과를 근거로 수행하는 의료 대응보다 효과가 떨어진다. 따라서 본 논문에서는 혈액검사로 도출되는 백혈구 수치를 GAN을 이용하여 예측하는 연구를 수행했다. 이는 환자의 생체신호와 정보를 바탕으로 예측되는 딥러닝 모델이며, 상대적으로 경험이 적어 의료적 대응에 어려움을 느끼는 의료진의 대응 보조 소프트웨어이다.

II. 본론

본 논문에서는 PhysioNet의 'Early Prediction of Sepsis from Clinical Data the PhysioNet Computing in Cardiology Challenge 2019'를 데이터 세트로 활용했다.[5] 이 데이터는 두 병원의 중환자실에서 획득된 패혈증 관련 정형 데이터이며, 본 연구에서는 해당 데이터의 41개 항목 중, 생체신호와 생체정보 및 환자 정보와 주석의 데이터를 추출하여 사용하였다. 그러나 생체신호와 환자 정보 및 주석을 제외한 생체정보 데이터는 결측값이 높다. 이는 생체정보가 여러 의료적 검사를 기반으로 도출되는 정량적 값으로 데이터 수집 단계에서 발생하는 검사 누락이 원인으로 생각된다.

Data Factor		정상(779,079)		패혈증(17,136)	
		Count	Rate(%)	Count	Rate(%)
Vital Sign	HR	60,084	7.712	1,105	6.448
	O2Sat	93,447	11.995	1,632	9.524
	Temp	511,616	65.669	11,698	68.266
	SBP	117,194	15.043	3,007	17.548
	MAP	79,477	10.201	1,381	8.059
	DBP	373,195	47.902	7,102	41.445
	Resp	75,922	9.745	1,336	7.796
Patient & Vital Info	Age	0	0	0	0
	pH	685,258	84.5	14,342	83.7
	HCO3	711,108	91.3	15,490	90.4
	PaCO2	705,999	90.6	14,928	87.1
	Creatinine	721,877	92.5	15,851	92.5
WBC	715,201	91.8	15,666	91.422	

표 1. 결측 데이터 통계표

표 1은 본 연구에서 사용하는 데이터와 결측값을 표기했다. 본 연구에서는 전체 데이터 796,215(정상 + 패혈증)개에서 모든 항목에 대한 결측값이 없는 케이스를 분류하여 사용하였다. 또한, 본 연구에서는 분류를 위해 Factor Filter를 사용했다. 이는 백혈구 수치를 제외한 모든 데이터 항목에 관하여 결측값을 조사하고 조건에 해당하는 데이터만을 추출하는 과정이다. 이 과정을 통하여, 전체 데이터 796,215개 중 5,904개의 데이터를 추출하여 백혈구 수치 예측모델에 사용한다. 백혈구 수치 예측모델에 활용되는 데이터는 정상 데이터 5,770개, 패혈증 데이터 134개이며, 정상 데이터의 경우에는 백혈구 수치를 포함한 다른 데이터 항목 1개가 결측된 상태를 포함한다. 반면에 패혈증 데이터의 경우에는 모든 데이터가 존재하는 상태로 결측된 데이터 항목이 없다.

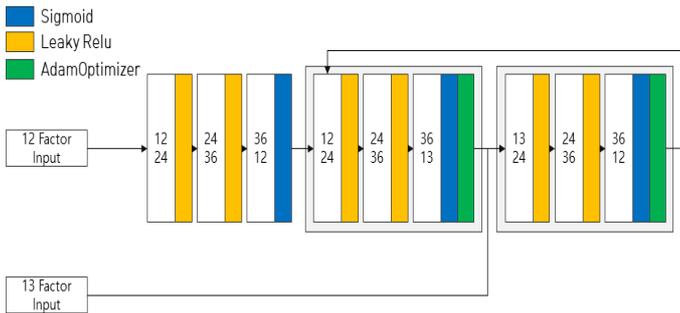


그림 1. 백혈구 수치 예측모델 구조

그림 1은 백혈구 수치 예측모델의 구조이다. 정형 데이터를 사용하는 본 연구에서는 복잡한 구조의 딥러닝 모델을 사용할 필요가 없다. 따라서 GAN의 구조는 기본적인 FC(Fully Connected)의 형태를 갖는다. 해당 모델은 모든 항목이 존재하는 데이터를 학습하였으며, 백혈구 수치가 결측된 데이터를 입력받아, 이를 예측한다. 본 모델은 기존 혈액검사의 결과가 도출되기 전까지의 의료적 대응 대기시간을 단축하고자 수행했던 연구로서, 모델의 실시간성을 고려하여 구성되었다. 따라서 최근 딥러닝 연구들에서 사용하는 높은 복잡도의 모델과 많은 차이가 있다.

본 연구의 검증은 백혈구가 결측된 원본데이터와 백혈구 수치를 예측한 보정 데이터를 섞어 인공지능 모델로 수행하였다. 검증에 사용한 인공지능 모델은 패혈증을 조기 검출하는 인공지능 모델이며, 예측된 백혈구 수치가 패혈증의 조기 검출 모델에 입력되었을 때, 발생하는 성능의 증감을 백혈구 수치의 검증으로 활용했다. 검증 결과는 표 2에서 볼 수 있다.

성능지표	원본 데이터군		혼합 데이터군	
	정상	패혈증	정상	패혈증
Specificity	0.64	0.92	1.00	0.95
Sensitivity	0.96	0.45	0.97	0.95
F1-Score	0.77	0.60	0.97	0.98
ACC	0.80	0.74	0.98	0.96

표 2. 백혈구 수치 예측을 통한 패혈증 예측 결과

성능지표는 Specificity, Sensitivity, F1-Score, ACC를 활용했다. 원본 데이터군의 경우에는 백혈구 수치 예측을 수행하지 않은 전체 데이터에서 무작위선별을 통해 수집한 200건의 데이터이다. 혼합 데이터군의 경우에는 마찬가지로 무작위선별을 통해 수집한 100건의 데이터와 백혈구 수치 예측모델을 통해 생성된 데이터 100건의 혼합 데이터이다.

표 2에서 볼 수 있듯이, 혼합 데이터군이 원본 데이터군보다 정상과 패혈증 환자에 대한 예측도가 높은 것으로 나타났다. 또한, 본 연구에서 고려하는 모든 성능지표도 높게 나타났다. 특히, 혼합 데이터군에서는 원본 데이터군보다 패혈증 환자에 대한 예측능력이 뛰어났다.

III. 결론

본 논문에서는 혈액검사를 통하여 도출할 수 있는 생체정보인 백혈구 수치를 GAN을 통하여 예측하는 연구를 수행했다. 그 결과로는 백혈구 수치 예측모델의 데이터를 혼합한 데이터군이 원본 데이터군보다 평균 20% 높은 예측 정확도를 보였다. 특히, 두 데이터군의 패혈증 환자 검출 성능의 차이가 큰 것으로 보았을 때, 백혈구 수치 예측모델의 결과가 패혈증을 예측하기 위한 데이터에서 중요한 요소로 활용되었을 것으로 사료된다. 또한, 두 데이터군에서 정상 환자의 분류에도 많은 영향을 끼친 것으로 보인다. 본 연구는 실시간성이 높은 백혈구 수치 예측모델의 연구로 의료적 대응이 긴급하게 필요한 응급실과 중환자실 환자들의 백혈구 수치를 예측하여, 의료 대응의 시간 감소 및 정밀도를 증가시킬 수 있을 것으로 기대된다. 향후 연구로는 생체신호와 생체정보 및 환자 정보의 입력 수준을 증가시켜, 정밀한 예측모델을 연구할 계획이다. 또한, 패혈증을 제외한 백혈구 수치가 적극적으로 활용되는 질병에 대하여, 예측모델을 연구할 계획이다.

ACKNOWLEDGMENT

이 논문은 2022년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 '바이오융복합기술 전문인력 양성사업'의 지원을 받아 수행된 연구임(No. P0017805).

참고 문헌

- [1] Smith, T. J., Bohlke, K., Lyman, G. H., Carson, K. R., Crawford, J., Cross, S. J., and Armitage, J. O. "Recommendations for the use of WBC growth factors: American Society of Clinical Oncology clinical practice guideline update," *Journal of Clinical Oncology*, 33(28), pp. 3199-3212, 2015.
- [2] Ge, Y. L., Liu, C. H., Rana, M. A., Zhu, X. Y., Wang, N., Xu, J., and Wang, H. Y. "Elevated Red Blood Cell Distribution Width Combined White Blood Cell in Peripheral Blood Routine Have a Better Sensitivity than CURB-65 Scores in Predicting ICU Admission and Mortality in Adult Community-Acquired Pneumonia Patients," *Clinical laboratory*, 65(3). 2019.
- [3] Goldstein, L. N., Wells, M., and Vincent-Lambert, C. "The cost of time: A randomised, controlled trial to assess the economic impact of upfront, point-of-care blood tests in the Emergency Centre," *African Journal of Emergency Medicine*, 9(2), pp. 57-63, 2019.
- [4] Hu, Z., Wang, H., Xie, J., Zhang, J., Li, H., Liu, S., and Huang, Y. "Burnout in ICU doctors and nurses in mainland China—a national cross-sectional study," *Journal of critical care*, vol. 62, pp. 265-270, 2021.
- [5] Reyna, M., Josef, C., Jeter, R., Shashikumar, S., Moody, B., Sharma, A., Nemat, S., and Clifford, G. "Early Prediction of Sepsis from Clinical Data - the PhysioNet Computing in Cardiology Challenge 2019," <https://physionet.org/content/challenge-2019/1.0.0/>, Aug. 2019.