



중환자실 폐렴 환자에 대한 시뮬레이션 기반 시계열 사망 마커 탐지

Simulation-Based Time Series Death Marker Detection for Pneumonia Patients in Intensive Care Unit

김수현^{1,2}, 이수현^{1,2,§}, 고가연^{1,2,§}, 안홍렬^{1,2,*}

Suhyun Kim^{1,2}, Suhyeon Lee^{1,2,§}, Gayoun Koh^{1,2,§} and Hongryul Ahn^{1,2,*}

1. 수원대학교 데이터과학부

2. 수원대학교 창업지원단 DS&ML 센터

§ : These authors contributed equally

서론

의료 빅데이터는 4차 산업혁명 빅데이터 분야의 핵심 응용 분야이다. 본 연구에서는 의료 빅데이터의 하나인 MIMIC III 데이터를 활용하여 중환자실 폐렴(pneumonia) 환자에 대한 시계열 사망 마커를 탐지하는 방법을 제안한다. 제안하는 방법으로 찾아낸 폐렴 사망 마커들로부터 폐렴의 사망 원인에 대한 새로운 단서를 발견하게 되면 폐렴의 사망률을 낮추는데 도움이 될 것으로 기대된다.

MIMIC-III

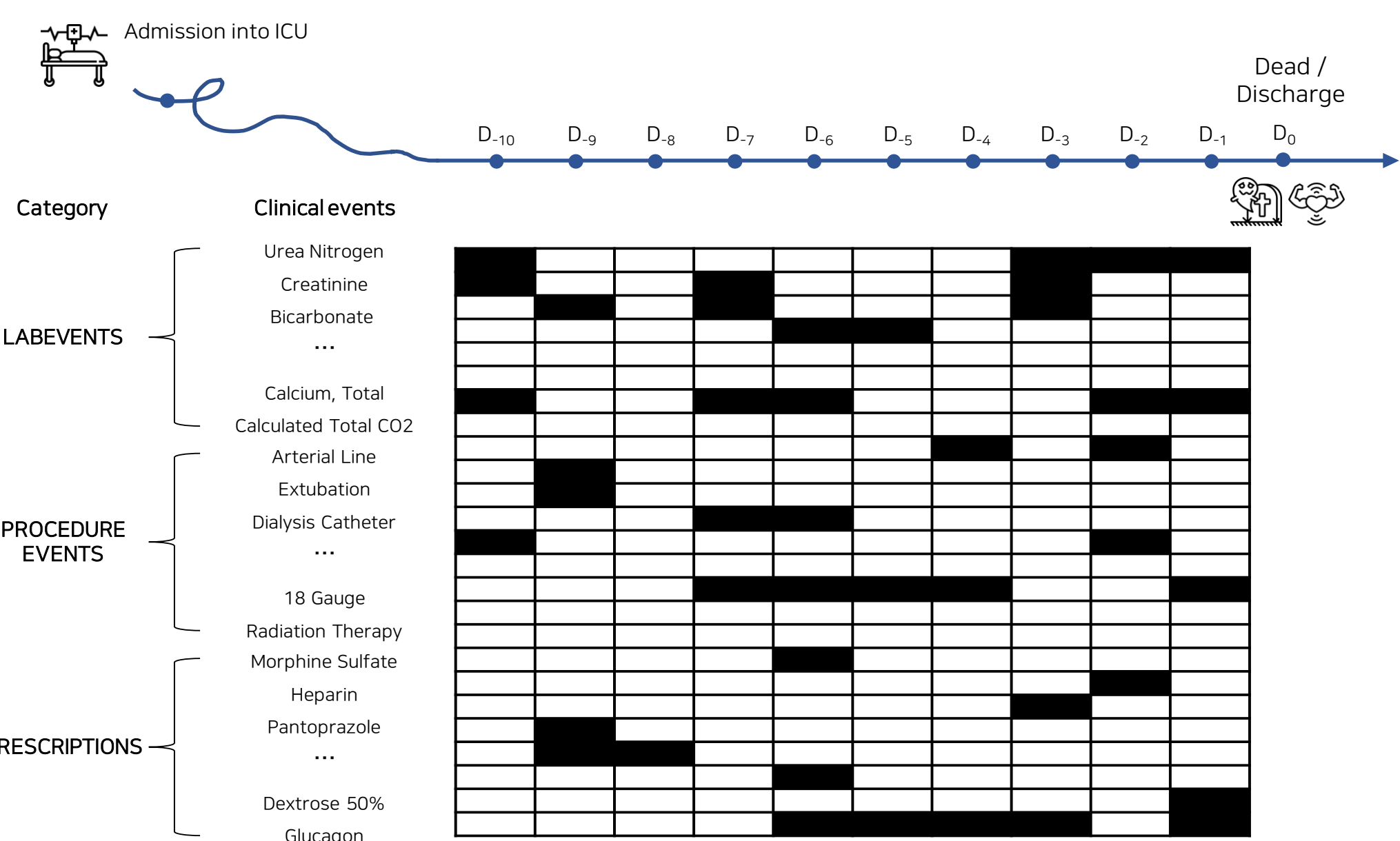
MIMIC-III(Medical Information Mart for Intensive Care III)은 2001년부터 2012년까지 Beth Israel Deaconess Medical Center의 중환자실에 머물렀던 4만 명 이상의 환자에 대해서 중환자실 입원 기간동안 발생한 의료 이벤트를 추적 및 기록한 시계열 전자 의무 기록(EMR : Electronic Medical Record) 데이터이다. [1] MIT Laboratory for Computational Physiology 연구실에서 익명화 및 구조화를 진행하고 2015년 공개되었다.

사망 마커

사망 마커란 해당 의료 이벤트가 발생했을 때, 폐렴 환자의 사망 경향이 높아지는 의료 이벤트를 의미한다.

문제 정의

1. 폐렴 환자 시계열 의료 데이터.

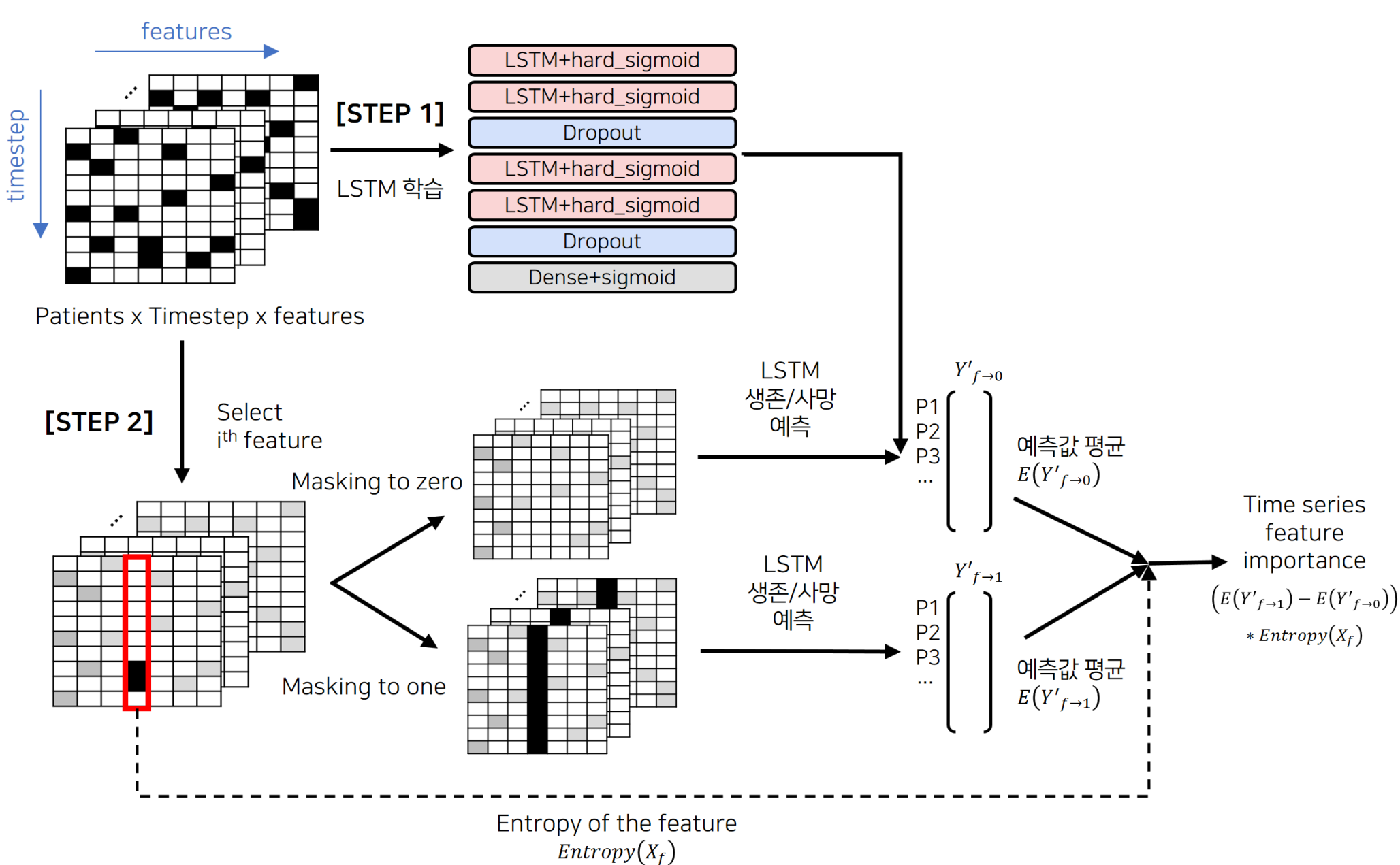


[그림 1] 한 환자의 10일치 의료 시계열 이벤트 데이터 예시. 가로축은 사망/퇴원일(D0)과 그 전 10일(D-10 ~ D-1)의 시점들이며, 세로축은 의료 이벤트의 종류이다. 데이터 가운데에 검은색/흰색 칸은 해당 의료 이벤트가 그 시점에서 발생/미발생 했음을 의미한다.

2. 폐렴 환자 사망 마커 탐지 문제.

폐렴 환자 시계열 데이터 X 에 존재하는 F 개의 의료 이벤트 $\{F_1, ..., F_F\}$ 에 대해서, 폐렴 환자가 사망에 이르게 하는 영향력에 대한 중요도 점수 $\{IS(F_1), ..., IS(F_F)\}$ 를 출력하는 문제이다(IS 는 importance score의 약자이다). 각 이벤트의 중요도 점수로부터 상위 n 개 이벤트를 선별하여 n 개의 사망 마커를, 하위 n 개 이벤트를 선별하여 n 개의 생존 마커를 최종적으로 산출하게 된다.

방법



[그림 2] 시뮬레이션 기반 시계열 사망 마커 점수 계산 방법.

[STEP 1] 시계열 인공지능 모델 학습. (그림 2의 상단부)

학습 시 기존의 장기 의존성 문제를 개선한 딥러닝 인공지능 모델인 LSTM(Long Short-Term Memory)을 사용하였다[2]. 손실 함수와 최적화 알고리즘으로 binary cross entropy와 adam을 사용하였으며, 최대 반복 학습 횟수(epoch)는 300회로 지정하고 과적합을 방지하기 위하여 early stopping을 적용하였다.

[STEP 2] 데이터 시뮬레이션을 통한 시계열 사망 마커 중요도 점수 계산. (그림 2의 하단부)

[1] 두 종류의 시뮬레이션 데이터 $X_{f \rightarrow 1}$, $X_{f \rightarrow 0}$ 구하기

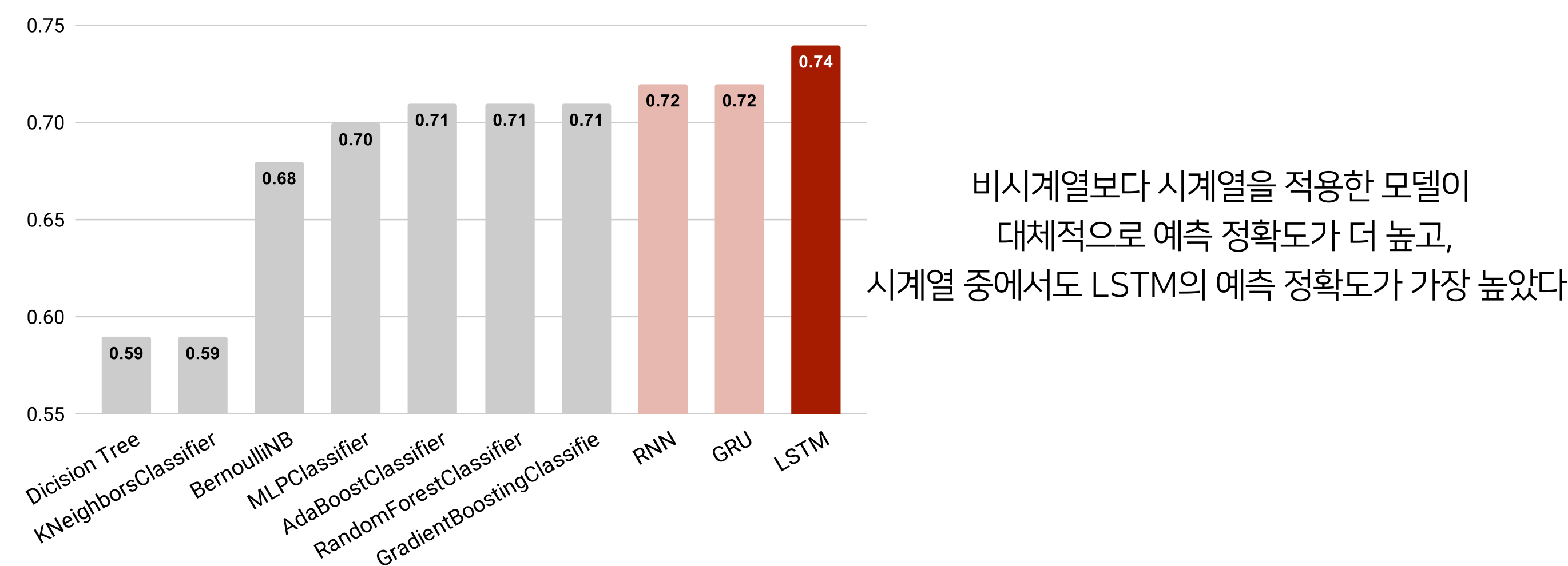
- $X_{f \rightarrow 1}$: 원본 데이터 X 에서 의료 이벤트 f 의 모든 환자의 모든 시점에 대한 값을 발생(1) 값으로 바꾼 발생 시뮬레이션 데이터이다.
- $X_{f \rightarrow 0}$: 원본 데이터 X 에서 의료 이벤트 f 의 모든 환자의 모든 시점에 대한 값을 미발생(0) 값으로 바꾼 미발생 시뮬레이션 데이터이다.

[2] 의료 이벤트 f 에 대한 사망 마커 점수 $IS(f)$ 구하기

$$IS(f) = (E(Y'_{f \rightarrow 1}) - E(Y'_{f \rightarrow 0})) * Entropy(X_f)$$

실험 및 결과

RESULT 1) 다른 모델과의 예측 정확도 비교.



[그림 3] 폐렴 환자 사망 예측에 대한 ROC-AUC 비교. 비시계열 모델 (회색) vs 시계열 모델 (연빨강) vs LSTM (진빨강).

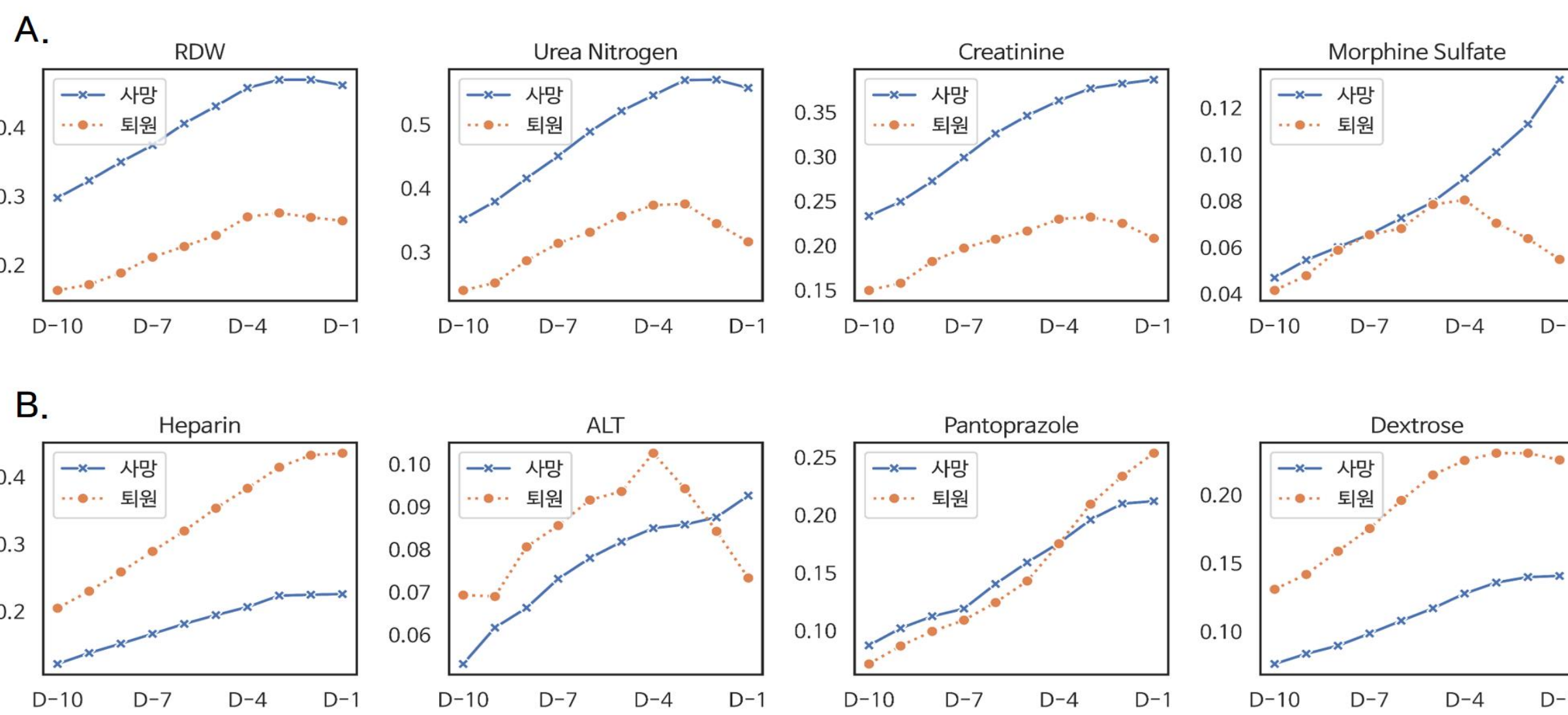
RESULT 2) 폐렴 환자의 사망/생존 의료 이벤트 top4.

[표 1] 사망 마커 TOP 4에 대한 폐렴 관련 문헌

이벤트	점수	폐렴 관련 문헌
적혈구 용적 분포폭치 (RDW)	0.108	RDW 수치가 정상범위인 13.8%보다 높으면 사망률이 약 1.7배 상승한다는 연구결과가 있다. [3]
혈액요소질소 (Blood Urea Nitrogen)	0.085	신장기능 이상을 측정하는 검사 항목. 신장이 건강하지 않은 폐렴 환자가 사망할 위험이 높다는 연구결과가 있다. [4]
크레아티닌 (Creatinine)	0.058	폐렴이 악화되었을 때, 그 사망군에서 비정상적인 크레아티닌 수치가 검출된 경우가 발견됐다는 연구결과가 있다. [5]
황산 모르핀 (Morphine Sulfate)	0.045	중환자실에서 가장 많이 사용되는 마약성 진통제. 중증 환자에 대한 진통제로 황산 모르핀이 추천되기도 한다. [6]

[표 2] 생존 마커 TOP 4에 대한 폐렴 관련 문헌

이벤트	점수	폐렴 관련 문헌
헤파린 (Heparin)	-0.054	급성 호흡 부전을 동반한 폐렴 환자에게 실시하는 저분자 헤파린요법은 사망률을 줄인다는 보고가 있다. [7]
알라닌 전달 효소 (ALT)	-0.040	특별한 관련 문헌을 찾을 수 없음.
판토프라졸 (Pantoprazole)	-0.037	특별한 관련 문헌을 찾을 수 없음.
포도당 (Dextrose)	-0.033	정상 범위 내에서 포도당 수치를 감소하고 유지하는 것은 성공적인 치료의 중요한 요소다. [8]



[그림 4] 사망 및 생존 환자에 대한 이벤트 발생 빈도 비율. (A) 사망 마커 (B) 생존 마커. 가로축은 사망/퇴원일(D0) 전 10일(D-10 ~ D-1)을, 세로축은 이벤트 발생 빈도 비율을 나타낸다.

결론

본 연구에서는 중환자실 폐렴 환자의 전자 의무 기록을 기반으로 생존 예측 모델을 구축하고 시뮬레이션에 기반한 시계열 사망 마커 탐지 기법을 제안하였다. MIMIC-III 데이터에서 폐렴 환자의 사망/생존 마커를 탐지하였을 때, 사망/생존에 관련된 의료 이벤트를 검출할 수 있음을 확인하였다.

참고문헌

- [1] Johnson A, Pollard T, Shen L, Lehman L, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific data. 2016 May;3(1):1-9.
- [2] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997;9(8):1735-1780.
- [3] Yoo KD, et al. Red blood cell distribution width as a predictor of mortality among patients regularly visiting the nephrology outpatient clinic. Scientific reports, 11(1):1-9. 2021.
- [4] Feng DY, et al. Elevated blood urea nitrogen-to-serum albumin ratio as a factor that negatively affects the mortality of patients with hospital-acquired pneumonia. Canadian Journal of Infectious Diseases and Medical Microbiology. 2019.
- [5] 장상민, et al. 응급실을 통해 중환자실에 입원한 패혈증 환자의 예후 예측. 대한응급의학회지, 27(4):306-312. 2016.
- [6] Shapiro BA, et al. Practice parameters for systemic intravenous analgesia and sedation for adult patients in the intensive care unit: an executive summary. Critical care medicine, 23:1596-1600. 1995.
- [7] 이미숙, et al. 성인 지역사회획득 폐렴 항생제 사용지침. Infection & Chemotherapy. 2018.
- [8] Kubisz A, et al. Elevated blood glucose level as a risk factor of hospital-acquired pneumonia among patients treated in the intensiv care unit (ICU). Przegląd Lekarski, 68(3):136-9. 2011.



중환자실 폐렴 환자에 대한 시뮬레이션 기반 시계열 사망 마커 탐지

Simulation-Based Time Series Death Marker Detection for Pneumonia Patients in Intensive Care Unit

김수현^{1,2}, 이수현^{1,2,§}, 고가연^{1,2,§}, 안홍렬^{1,2,*}

Suhyun Kim^{1,2}, Suhyeon Lee^{1,2,§}, Gayoun Koh^{1,2,§} and Hongryul Ahn^{1,2,*}

1. 수원대학교 데이터과학부

2. 수원대학교 창업지원단 DS&ML 센터

§ : These authors contributed equally

서론

의료 빅데이터는 4차 산업혁명 빅데이터 분야의 핵심 응용 분야이다. 본 연구에서는 의료 빅데이터의 하나인 MIMIC III 데이터를 활용하여 중환자실 폐렴(pneumonia) 환자에 대한 시계열 사망 마커를 탐지하는 방법을 제안한다. 제안하는 방법으로 찾아낸 폐렴 사망 마커*들로부터 폐렴의 사망 원인에 대한 새로운 단서를 발견하게 되면 폐렴의 사망률을 낮추는데 도움이 될 것으로 기대된다.

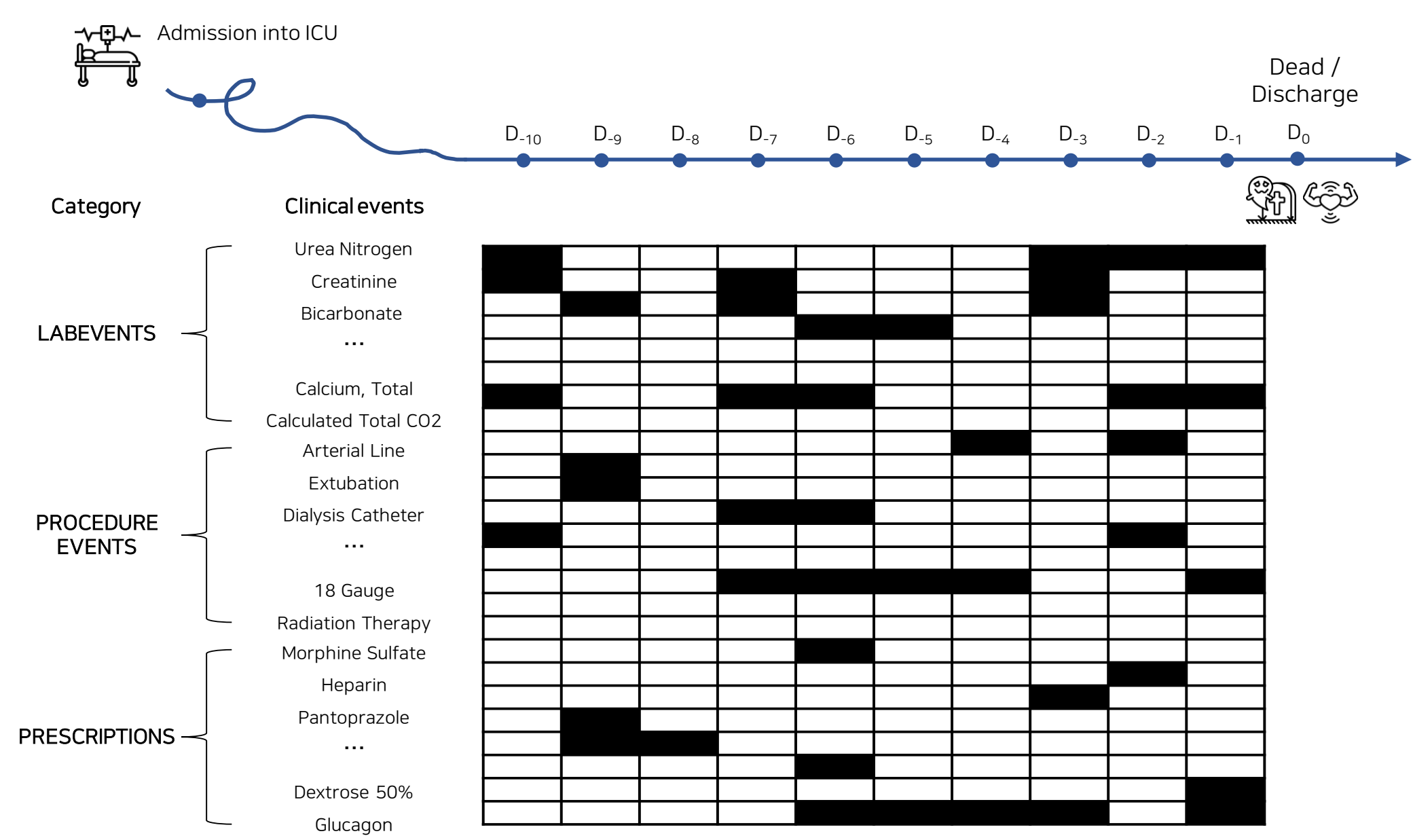
MIMIC-III

MIMIC-III (Medical Information Mart for Intensive Care III)은 2001년부터 2012년까지 Beth Israel Deaconess Medical Center의 중환자실에 머물렀던 4만 명 이상의 환자에 대해서 중환자실 입원 기간동안 발생한 의료 이벤트를 추적 및 기록한 시계열 전자 의무 기록(EMR : Electronic Medical Record) 데이터이다. [1] MIT Laboratory for Computational Physiology 연구실에서 익명화 및 구조화를 진행하고 2015년 공개되었다.

* 사망 마커란 해당 의료 이벤트가 발생했을 때, 폐렴 환자의 사망 경향이 높아지는 의료 이벤트를 의미한다.

문제 정의

1. 폐렴 환자 시계열 의료 데이터.

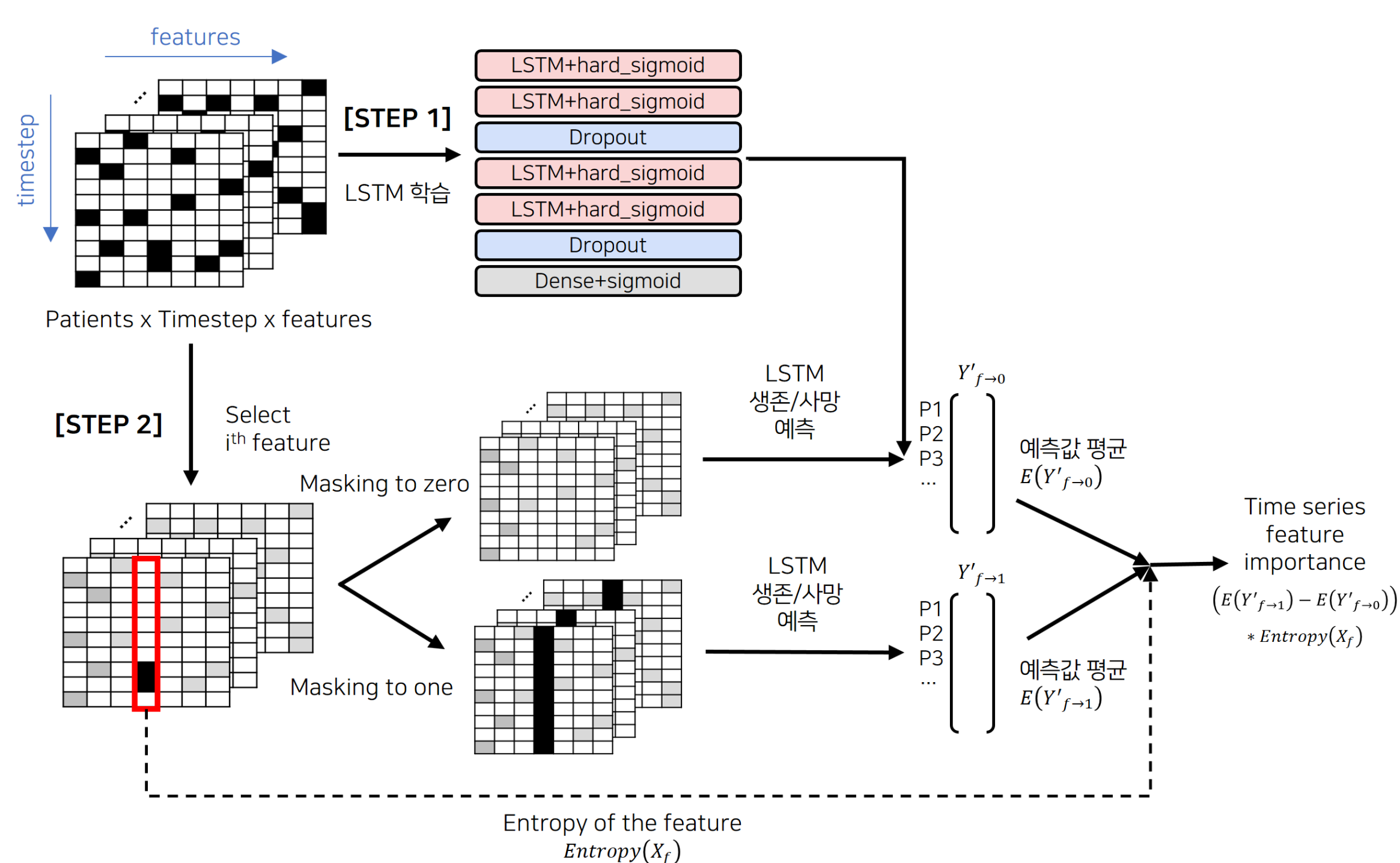


[그림 1] 한 환자의 10일치 의료 시계열 이벤트 데이터 예시. 가로축은 사망/퇴원일(D0)과 그 전 10일(D-10 ~ D-1)의 시점들이며, 세로축은 의료 이벤트의 종류이다. 데이터 가운데에 검은색/흰색 칸은 해당 의료 이벤트가 그 시점에서 발생/미발생 했음을 의미한다.

2. 폐렴 환자 사망 마커 탐지 문제.

폐렴 환자 시계열 데이터 X 에 존재하는 F 개의 의료 이벤트 $\{F_1, \dots, F_F\}$ 에 대해서, 폐렴 환자가 사망에 이르게 하는 영향력에 대한 중요도 점수 $\{IS(F_1), \dots, IS(F_F)\}$ 를 출력하는 문제이다(IS 는 importance score의 약자이다). 각 이벤트의 중요도 점수로부터 상위 n 개 이벤트를 선별하여 n 개의 사망 마커를, 하위 n 개 이벤트를 선별하여 n 개의 생존 마커를 최종적으로 산출하게 된다.

방법



[그림 2] 시뮬레이션 기반 시계열 사망 마커 점수 계산 방법.

[STEP 1] 시계열 인공지능 모델 학습. (그림 2의 상단부)

학습 시 기존의 장기 의존성 문제를 개선한 딥러닝 인공지능 모델인 LSTM(Long Short-Term Memory)을 사용하였다[2]. 손실 함수와 최적화 알고리즘으로 binary cross entropy와 adam을 사용하였으며, 최대 반복 학습 횟수(epoch)는 300회로 지정하고 과적합을 방지하기 위하여 early stopping을 적용하였다.

[STEP 2] 데이터 시뮬레이션을 통한 시계열 사망 마커 중요도 점수 계산. (그림 2의 하단부)

[1] 두 종류의 시뮬레이션 데이터 $X_{f \rightarrow 1}$, $X_{f \rightarrow 0}$ 구하기

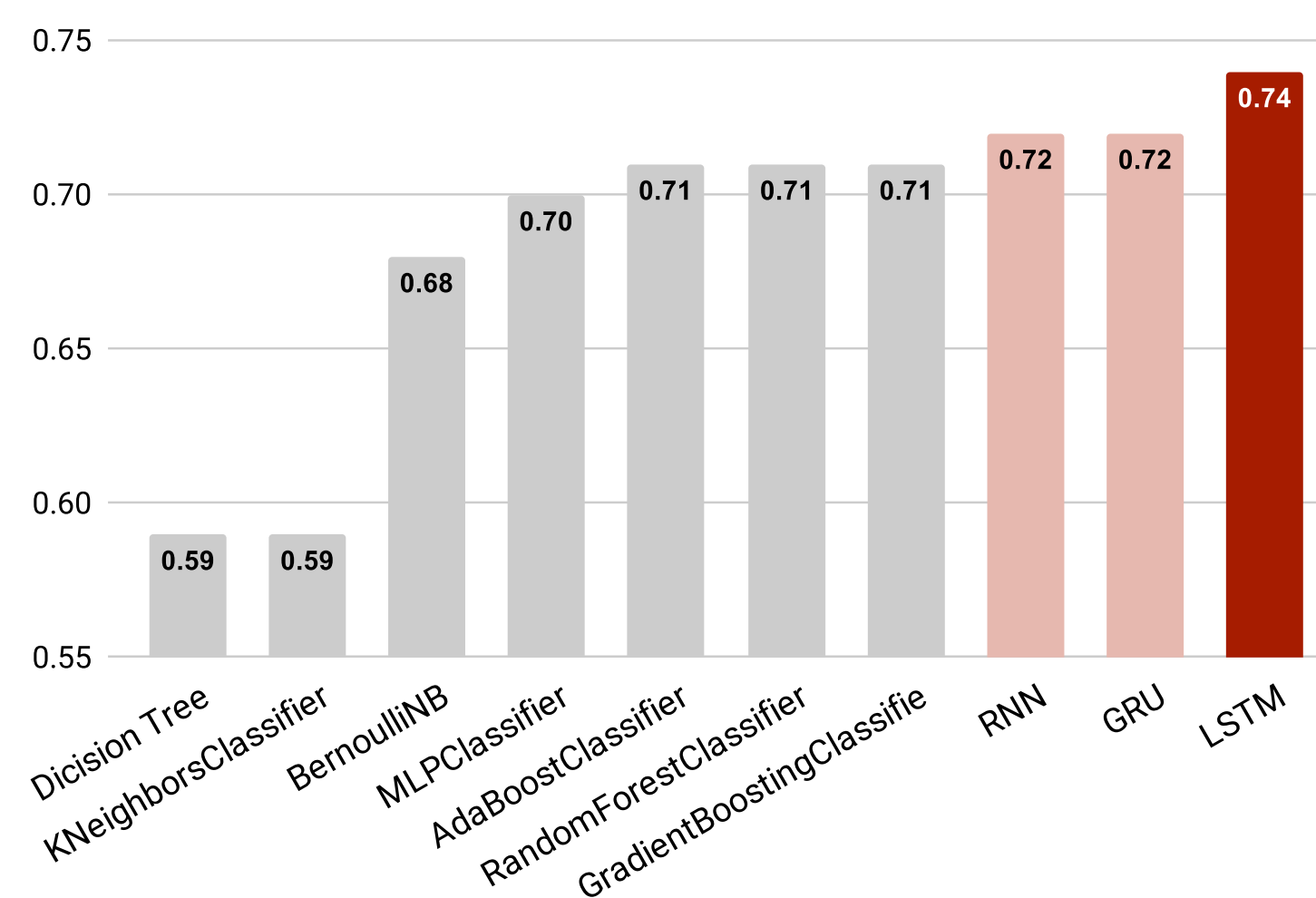
- $X_{f \rightarrow 1}$: 원본 데이터 X 에서 의료 이벤트 f 의 모든 환자의 모든 시점에 대한 값을 발생(1) 값으로 바꾼 발생 시뮬레이션 데이터이다.
- $X_{f \rightarrow 0}$: 원본 데이터 X 에서 의료 이벤트 f 의 모든 환자의 모든 시점에 대한 값을 미발생(0) 값으로 바꾼 미발생 시뮬레이션 데이터이다.

[2] 의료 이벤트 f 에 대한 사망 마커 점수 $IS(f)$ 구하기

$$IS(f) = (E(Y'_{f \rightarrow 1}) - E(Y'_{f \rightarrow 0})) * Entropy(X_f)$$

실험 및 결과

RESULT 1) 다른 모델과의 예측 정확도 비교.



비시계열보다 시계열을 적용한 모델이 대체적으로 예측 정확도가 더 높고, 시계열 중에서도 LSTM의 예측 정확도가 가장 높았다.

[그림 3] 폐렴 환자 사망 예측에 대한 ROC-AUC 비교. 비시계열 모델 (회색) vs 시계열 모델 (연빨강) vs LSTM (진빨강).

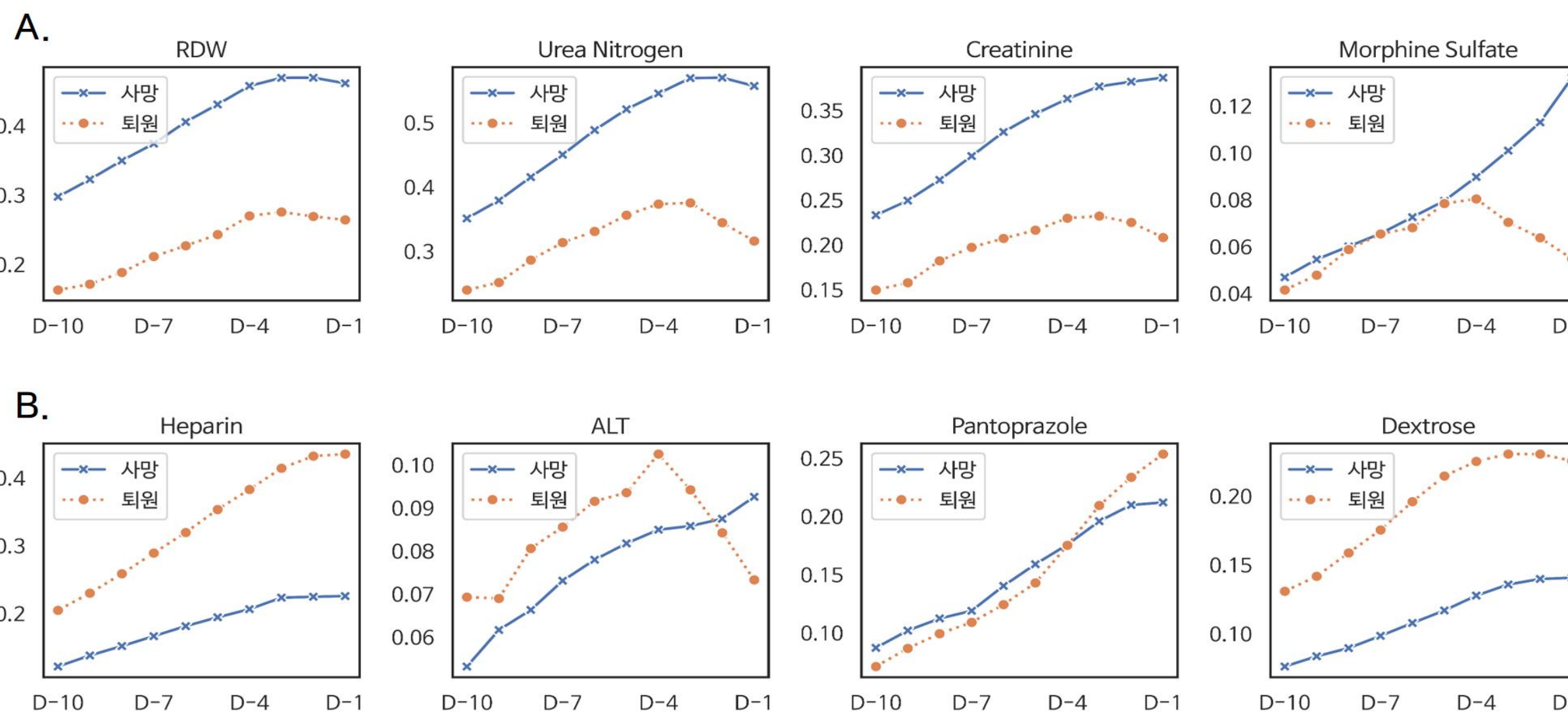
RESULT 2) 폐렴 환자의 사망/생존 의료 이벤트 top4.

[표 1] 사망 마커 TOP 4에 대한 폐렴 관련 문헌

이벤트	점수	폐렴 관련 문헌
적혈구 용적 분포폭치 (RDW)	0.108	RDW 수치가 정상범위인 13.8%보다 높으면 사망률이 약 1.7배 상승한다는 연구결과가 있다. [3]
혈액요소질소 (Blood Urea Nitrogen)	0.085	신장기능이상을 측정하는 검사 항목. 신장이 건강하지 않은 폐렴 환자가 사망할 위험이 높다는 연구결과가 있다. [4]
크레아티닌 (Creatinine)	0.058	폐렴이 악화되었을 때, 그 사망군에서 비정상적인 크레아티닌 수치가 검출된 경우가 발견됐다는 연구결과가 있다. [5]
황산 모르핀 (Morphine Sulfate)	0.045	중환자실에서 가장 많이 사용되는 마약성 진통제. 중증 환자에 대한 진통제로 황산 모르핀이 추천되기도 한다. [6]

[표 2] 생존 마커 TOP 4에 대한 폐렴 관련 문헌

이벤트	점수	폐렴 관련 문헌
헤파린 (Heparin)	-0.054	급성 호흡 부전을 동반한 폐렴 환자에게 실시하는 저분자 헤파린요법은 사망률을 줄인다는 보고가 있다. [7]
알라닌 전담 효소 (ALT)	-0.040	특별한 관련 문헌을 찾을 수 없음.
판토프라졸 (Pantoprazole)	-0.037	특별한 관련 문헌을 찾을 수 없음.
포도당 (Dextrose)	-0.033	정상 범위 내에서 포도당 수치를 감시하고 유지하는 것은 성공적인 치료의 중요한 요소다. [8]



[그림 4] 사망 및 생존 환자에 대한 이벤트 발생 빈도 비율. (A) 사망 마커 (B) 생존 마커. 가로축은 사망/퇴원일(D0) 전 10일(D-10 ~ D-1)을, 세로축은 이벤트 발생 빈도 비율을 나타낸다.

결론

본 연구에서는 중환자실 폐렴 환자의 전자 의무 기록을 기반으로 생존 예측 모델을 구축하고 시뮬레이션에 기반한 시계열 사망 마커 탐지 기법을 제안하였다. MIMIC-III 데이터에서 폐렴 환자의 사망/생존 마커를 탐지하였을 때, 사망/생존에 관련된 의료 이벤트를 검출할 수 있음을 확인하였다.

참고문헌

- [1] Johnson A, Pollard T, Shen L, Lehman L, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific data. 2016 May;3(1):1-9.
- [2] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997;9(8):1735-1780.
- [3] Yoo KD, et al. Red blood cell distribution width as a predictor of mortality among patients regularly visiting the nephrology outpatient clinic. Scientific reports, 11(1):1-9. 2021.
- [4] Feng DY, et al. Elevated blood urea nitrogen-to-serum albumin ratio as a factor that negatively affects the mortality of patients with hospital-acquired pneumonia. Canadian Journal of Infectious Diseases and Medical Microbiology. 2019.
- [5] 장상민, et al. 응급실을 통해 중환자실에 입원한 패혈증 환자의 예후 예측. 대한응급의학회지, 27(4):306-312. 2016.
- [6] Shapiro BA, et al. Practice parameters for systemic intravenous analgesia and sedation for adult patients in the intensive care unit: an executive summary. Critical care medicine, 23:1596-1600. 1995.
- [7] 이미숙, et al. 성인 지역사회획득 폐렴 항생제 사용지침. Infection & Chemotherapy. 2018.
- [8] Kubisz A, et al. Elevated blood glucose level as a risk factor of hospital-acquired pneumonia among patients treated in the intensiv care unit (ICU). Przegląd Lekarski, 68(3):136-9. 2011.