



# 중환자실 패혈증 사망 예측을 위한 다중 조합 시간적 트랜스포머 모델

Multi-Combination Temporal Interval Transformer Model for Predicting Mortality in ICU Sepsis Patients

조희연<sup>1,2</sup>, 안홍렬<sup>1,2,\*</sup>

Huiyeon Joe<sup>1,2</sup> and Hongryul Ahn<sup>1,2,\*</sup>  
1. 수원대학교 데이터과학부  
2. 수원대학교 창업지원단 DS&ML 센터

## 서 론

기존의 사망 예측 딥러닝 모델들은 주로 정적인 임상 정보를 기반으로 하거나, 특정 시간동안(예: 하루동안)에 발생된 데이터를 사용하거나, 일부 시간 동안만의 데이터를 사용함으로써 환자의 시간에 따른 상태 변화를 충분히 반영하지 못하는 한계가 있었다[1,2]. 특히, 중환자실에서는 환자의 생체 신호와 임상 지표가 시간에 따라 급격하게 변동하므로, 이러한 시간적 패턴을 효과적으로 모델링하는 것이 중요하다[3].

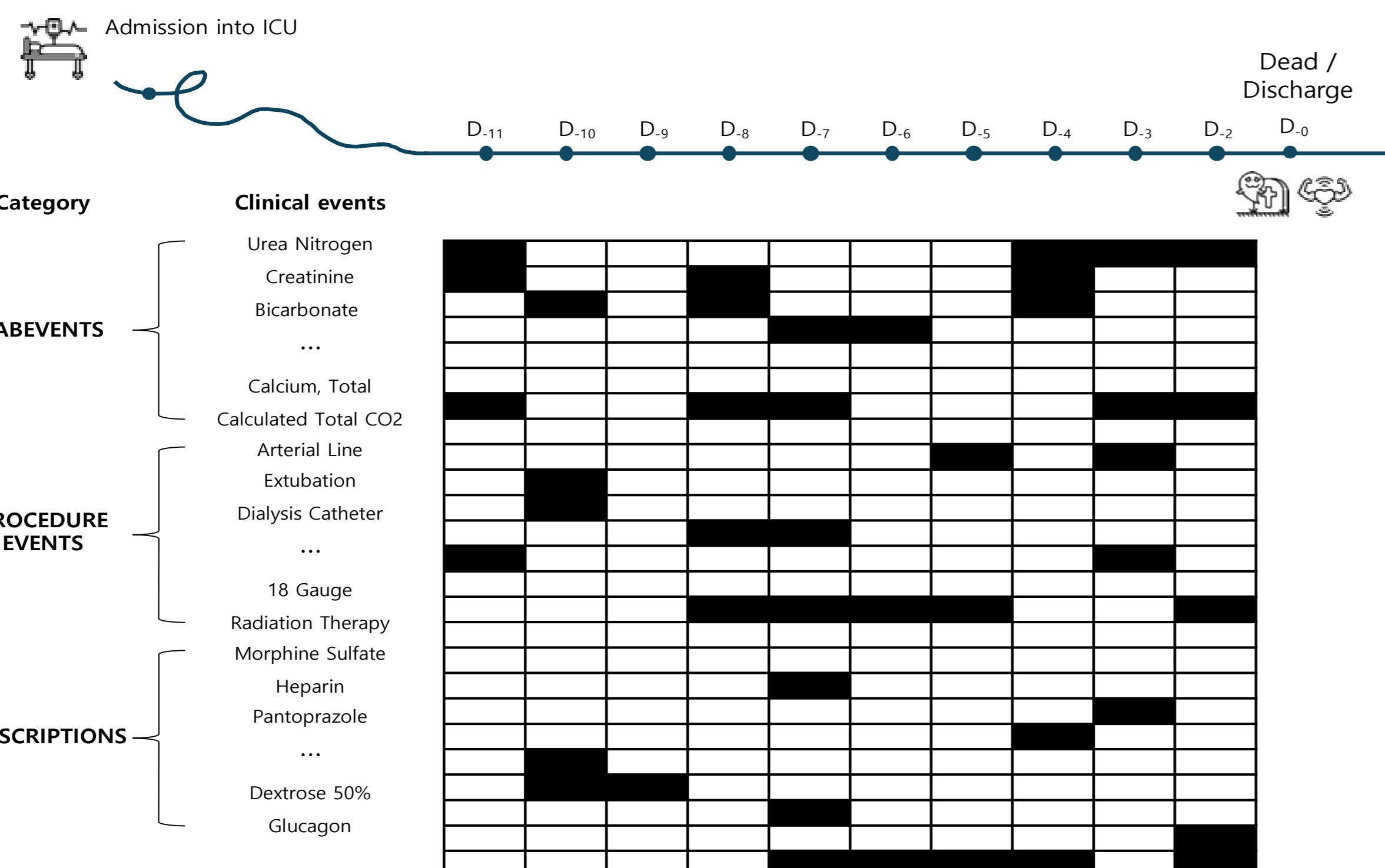
본 연구에서는 중환자실 패혈증 환자의 사망 예측을 위한 다중 조합 시간적 트랜스포머 모델을 제안하고 전자의무기록 데이터에서 해당 모델을 실험하여 검증한다.

### MIMIC-IV

MIMIC-IV(Medical Information Mart for Intensive Care IV)는 미국 Beth Israel Deaconess Medical Center에서 2008년부터 2019년까지 65,366명의 환자의 94,458건의 중환자실 입원 기록을 비식별화하고 연구 목적으로 공개한 대표적인 전자건강기록 데이터베이스이다[4].

## 시계열 데이터셋 구축

패혈증 환자 사망 예측 시계열 의료 데이터셋.



[그림 1] 한 환자의 24시간적 시계열 행렬 데이터 예시.

$D_0$ 는 사망일이며, 24시간적 시계열 데이터는  $D_{-12}$ 에서  $D_0$ 까지의 데이터를 설명변수 데이터로 사용한다. 이 때, 설명 변수 데이터로 사용되는 것은 실험(lab), 시술(procedure), 처방(prescription)에 대한 이진화 표현 데이터이다. 그림은 기존 논문[5]의 그림에서 수정하였다.

패혈증 환자의 중환자실 입원일에( $D_0$ ) 대해서, 퇴원일의 전전날( $D_{-2}$ )을 기준일로 설정하였다. 이때, 인공지능 모델은  $D_{-2}$ 의 이전의 시계열 설명변수 데이터를 사용하여,  $D_0$ 의 사망 또는 회복을 예측하는 문제를 해결한다. 이러한 시계열 설명 변수 데이터셋 구축을 위해서, 기준일로부터 6시간, 12시간, 24시간, 48시간 간격의 시계열 데이터를 구성하였다.

- 6시간 간격(6h interval) 데이터셋의 시계열은 4.5일전부터 2일전까지의 10개 시간적의 시계열  $T_6 = [(D_{-4.5}, D_{-4.25}), (D_{-4.25}, D_{-4}), (D_{-4}, D_{-3.75}), \dots, (D_{-3.25}, D_{-2})]$  이다.
- 12시간 간격(12h interval) 데이터셋의 시계열은 7일전부터 2일전까지의 10개 시간적의 시계열  $T_{12} = [(D_{-7}, D_{-6.5}), (D_{-6.5}, D_{-6}), (D_{-6}, D_{-5.5}), \dots, (D_{-3.5}, D_{-2})]$  이다.
- 24시간 간격(24h interval) 데이터셋의 시계열은 12일전부터 2일전까지의 10개 시간적의 시계열  $T_{24} = [(D_{-12}, D_{-11}), (D_{-11}, D_{-10}), (D_{-10}, D_{-9}), \dots, (D_{-3}, D_{-2})]$  이다.
- 48시간 간격(6h interval) 데이터셋의 시계열은 22일전부터 2일전까지의 10개 시간적의 시계열  $T_{48} = [(D_{-22}, D_{-20}), (D_{-20}, D_{-18}), (D_{-18}, D_{-16}), \dots, (D_{-4}, D_{-2})]$  이다.

위에 시간적에 대해서, 실험(lab), 시술(procedure), 처방(prescriptions)에서 의료 이벤트가 발생했을 때를 1, 의료 이벤트가 발생하지 않았을 때를 0으로 이진화 표현하여, 시계열 설명변수 데이터셋을 구축하였다.

## 방 법

본 연구에서는 6시간적, 12시간적, 24시간적, 48시간적의 다중 시간적 데이터를 사용하여 환자의 사망 여부를 예측하는 다중 조합 시간적 트랜스포머 모델을 제안하였다[그림 2].

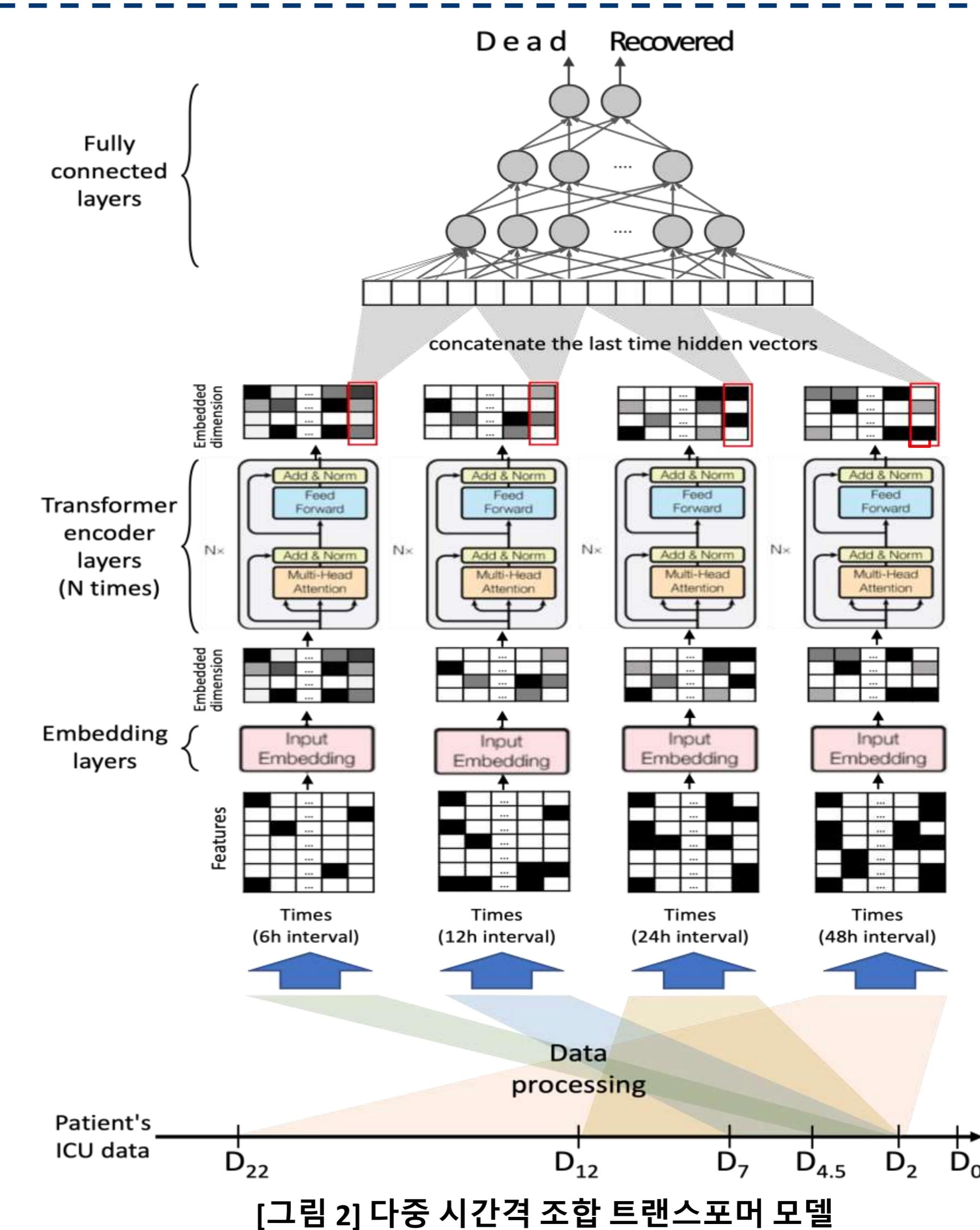
이 모델은 데이터 전처리를 통해 생성된 4종류의 시간적 행렬 데이터  $x_{6h}, x_{12h}, x_{24h}, x_{48h}$ 를 입력으로 받는다.

각 입력 데이터는 임베딩 층(E)과 트랜스포머 인코더 2개 층( $T^1, T^2$ )을 거쳐 히든 행렬 데이터로 변환된다.

$$\begin{aligned} h_{6h} &= T^2(T^1(E(x_{6h}))) \\ h_{12h} &= T^2(T^1(E(x_{12h}))) \\ h_{24h} &= T^2(T^1(E(x_{24h}))) \\ h_{48h} &= T^2(T^1(E(x_{48h}))) \end{aligned}$$

각 히든 행렬 데이터에서 마지막 시점의 히든 벡터는  $\oplus$  (concatenate) 연결되고, 완전 연결(fully connected) 층(F)을 거쳐 사망/회복 분류 예측으로 출력된다.

$$out = F(h_{6h}[-1] \oplus h_{12h}[-1] \oplus h_{24h}[-1] \oplus h_{48h}[-1])$$



## 결과

[표 1] 시간 간격별 평균 정확도 비교 결과. 진한 글씨는 정확도가 젤 높음을 의미.

모델	Precision	Recall	F1	Accuracy	ROC-AUC
<b>M<sub>6h</sub></b>	0.393	0.740	0.503	0.733	0.822
<b>M<sub>12h</sub></b>	0.433	0.730	0.531	0.764	0.841
<b>M<sub>24h</sub></b>	0.454	0.719	0.544	0.784	0.839
<b>M<sub>48h</sub></b>	0.366	<b>0.778</b>	0.488	0.703	0.829
<b>M<sub>6h+12h</sub></b>	0.463	0.753	0.561	0.783	0.854
<b>M<sub>6h+12h+24h</sub></b>	0.498	0.725	0.581	0.809	0.859
<b>M<sub>6h+12h+24h+48h</sub></b>	<b>0.508</b>	0.729	<b>0.587</b>	<b>0.811</b>	<b>0.863</b>

우리는 다중 조합 시간적 트랜스포머 모델의 성능을 검증하기 위해서, 단일 시간적 모델과 비교 실험을 수행하였다. 단일 시간적 모델은 단일 시간적 데이터만 사용하여 패혈증 환자의 사망을 예측하는 모델이다. 예를 들어,  $M_{6h}$  모델은 6시간 간격 데이터만 사용하는 모델을 말하며,  $M_{6h+12h}$ 는 6시간 간격 데이터와 12시간 간격 데이터를 사용하는 모델을 말한다.

단일 시간적 모델과 조합 시간적 모델을 비교하였을 때, 4종류의 모든 시간적을 다 사용한  $M_{6h+12h+24h+48h}$  모델이 Recall을 제외한 Precision, F1, Accuracy, ROC-AUC 지표에서 가장 높은 정확도를 보여주었다(표 1). 또한, 단일 시간적 모델( $M_{6h}, M_{12h}, M_{24h}, M_{48h}$ ), 2개 조합 시간적 모델( $M_{6h+12h}$ ), 3개 조합 시간적 모델( $M_{6h+12h+24h}$ ), 4개 조합 시간적 모델  $M_{6h+12h+24h+48h}$ , 순서로 F1, Accuracy, ROC-AUC 정확도가 상승하는 경향성을 보임으로써, 다양한 시간적 패턴을 통합할수록 생존 예측 정확도가 향상하는데 효과가 있음을 확인하였다.

## 결 론

본 연구는 중환자실(ICU)에서 패혈증 환자의 사망을 예측하기 위해 6시간, 12시간, 24시간, 48시간 간격의 다중 시간적 패턴을 통합한 트랜스포머 모델을 제안하고, 대규모 전자의무기록(EHR) 데이터베이스인 MIMIC-IV를 활용하여 검증하였다. 기존 모델들이 단일 시간 범위나 정적인 임상 정보에 의존하여 환자의 상태 변화를 충분히 반영하지 못하는 한계를 보완하여, 제안된 모델이 사망 예측 정확도를 크게 향상시킬 수 있음을 확인하였다. 이는 환자의 예후 예측에 중요한 인사이트를 제공한다.

## 참 고 문 헌

- [1] Johnson, A. E. W., Ghassemi, M. M., Nemati, S., et al. "Machine Learning and Decision Support in Critical Care." Proceedings of the IEEE , 104(2), 444-466, 2016.
- [2] Desautels, T., Calvert, J., Hoffman, J., et al. "Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach." JMIR Medical Informatics , 4(3), 2016.
- [3] Lehman, L. W., Adams, R. P., Mayaud, L., et al. "A Physiological Time Series Dynamics-Based Approach to Patient Monitoring and Outcome Prediction." IEEE Journal of Biomedical and Health Informatics, 19(3), 1068-1076, 2015.
- [4] Johnson, A. E. W., Pollard, T. J., Shen, L., et al. "MIMIC-IV, a Freely Accessible Electronic Health Record Dataset." Nature Scientific Data, 7, 2020.
- [5] 김수현, et al. "중환자실 패혈증 환자에 대한 시뮬레이션 기반 시계열 사망 마커 탐지" 한국정보과학회 학술발표논문집, pp.1952- 1954, 2022.