이상민 윤대관 윤태영 김태인 이상우

병원 전자 의무 기록 데이터를 활용한 **패혈증 환자 분석**

도파민팀

목차

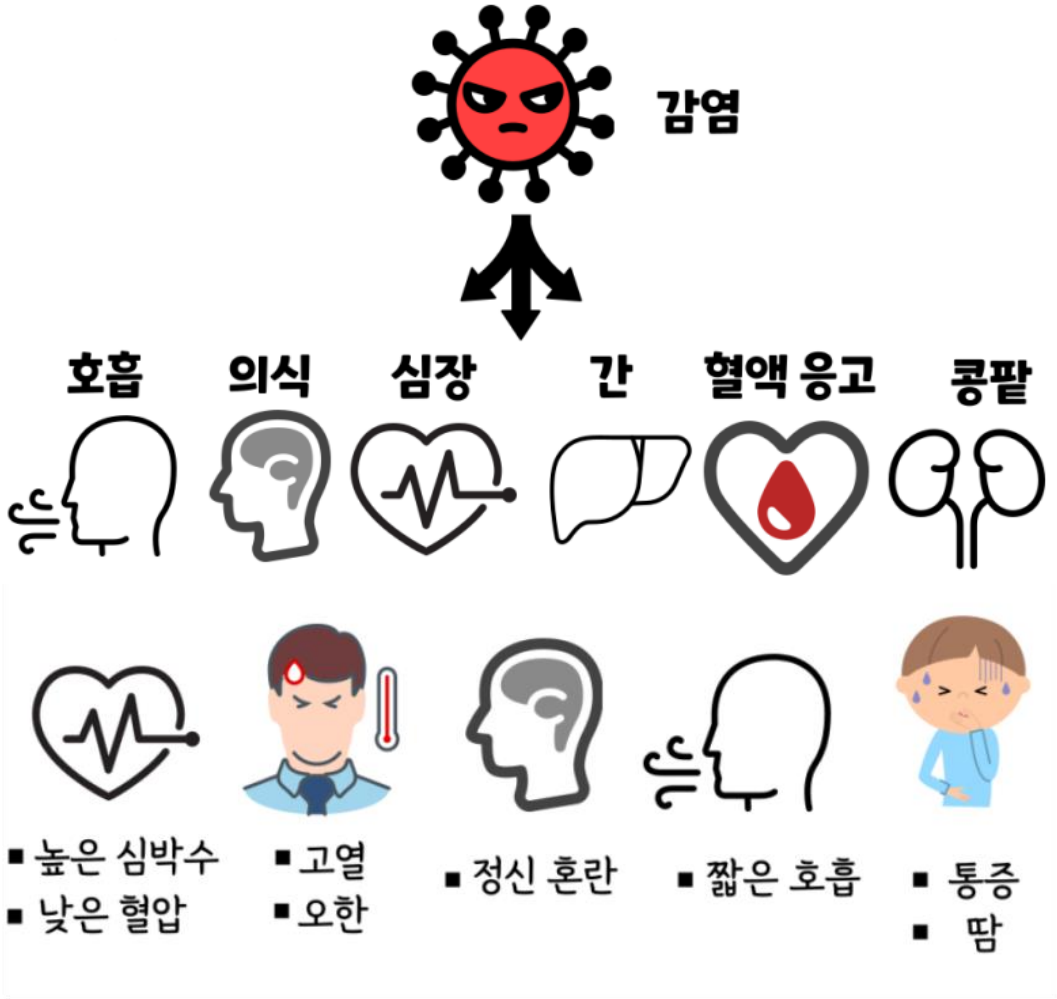
Part 1.

분석동기 및 데이터 설명

패혈증

혈액이 인체에 침입한 세균에 감염됨으로써 나타나는 전신성 염증반응 증후군

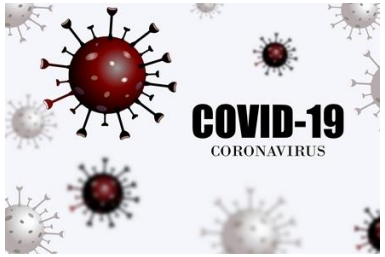
시간이 지날수록 사망률이 급격히 상승하며 많게는 50%까지 사망확률이 높은 질병



패혈증의 증가 요인

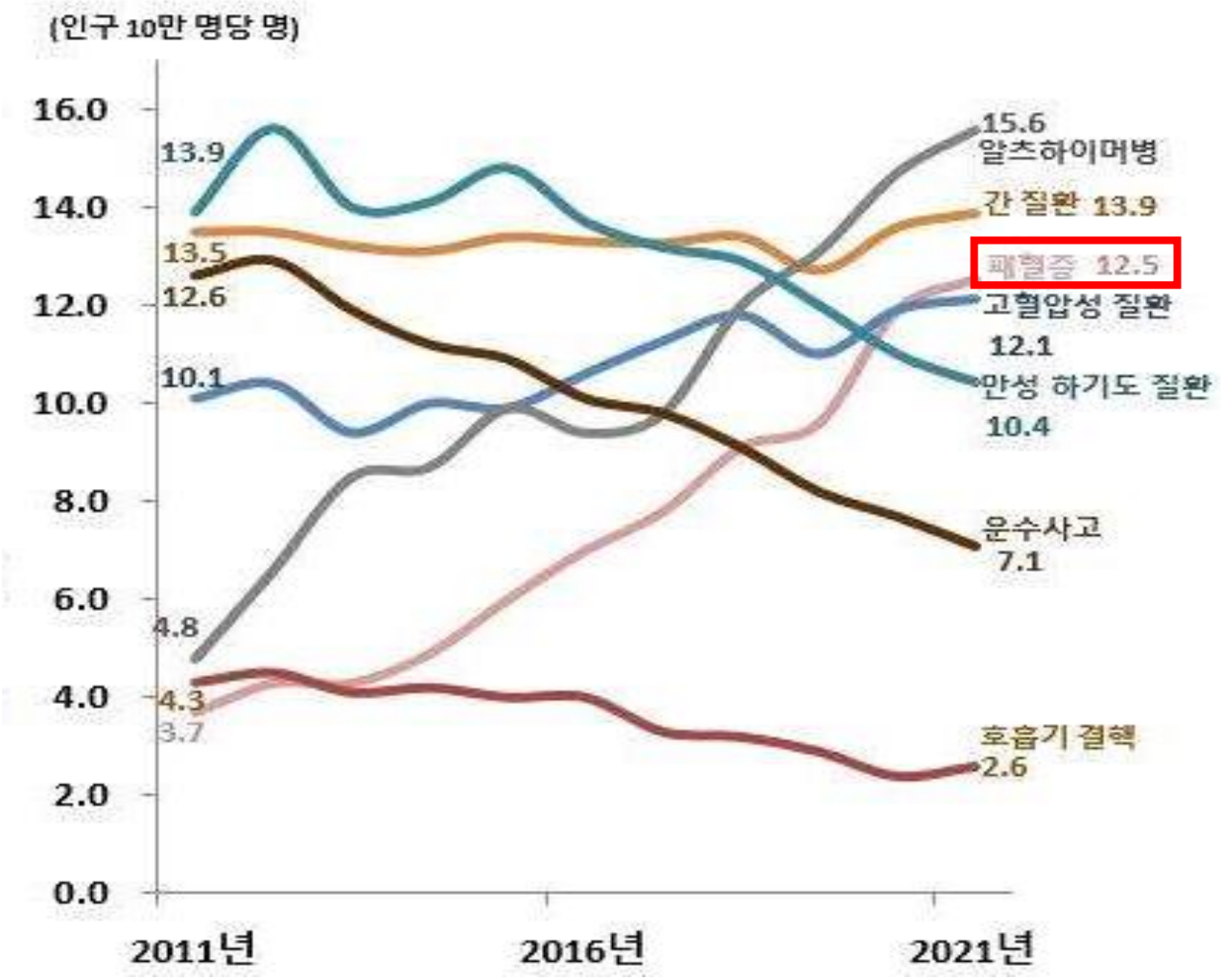
고령인구 증가
10년간 11% -> 46% 증가

당뇨환자 증가
8년간 330만 -> 600만 증가



코로나로 면역력 저하
2018~2021년 사이 급격한 변화

10년간 패혈증 사망률 4배 증가



패혈증에 대한 관심, 시장의 증가

사망 원인에도 드리운 고령화 그림자...패혈증, 사인 10위로 올라서

지난해 한국인이 사망한 10대 원인 가운데 처음으로 패혈증이 직접적인 사망 사유로 이름을 올렸다. 패혈증 은 노인이나 만성 질환자에게 발병하는 병으로, 패혈증 사망 증가는 고령 사회로 진입한 증거라는 분석이 나온다.

입력 2021.09.28 12:00

패혈증 진단 시장 크기, 공유, 동향, 기회 분석 예측 보고서 2030년까지

Tushar Jane
Search Engine Optimization Executive at Vantage Market Research

+ 팔로우

발행일: 2023년 5월 24일

글로벌 **패혈증 진단 시장은** 2022년에 USD 5억 2,950만 달러로 평가되었으며 예측 기간 동안 연평균 성장률(CAGR) 9.2%로 2030년까지 USD 9억 8,040만 달러에 이를 것으로 예상됩니다.

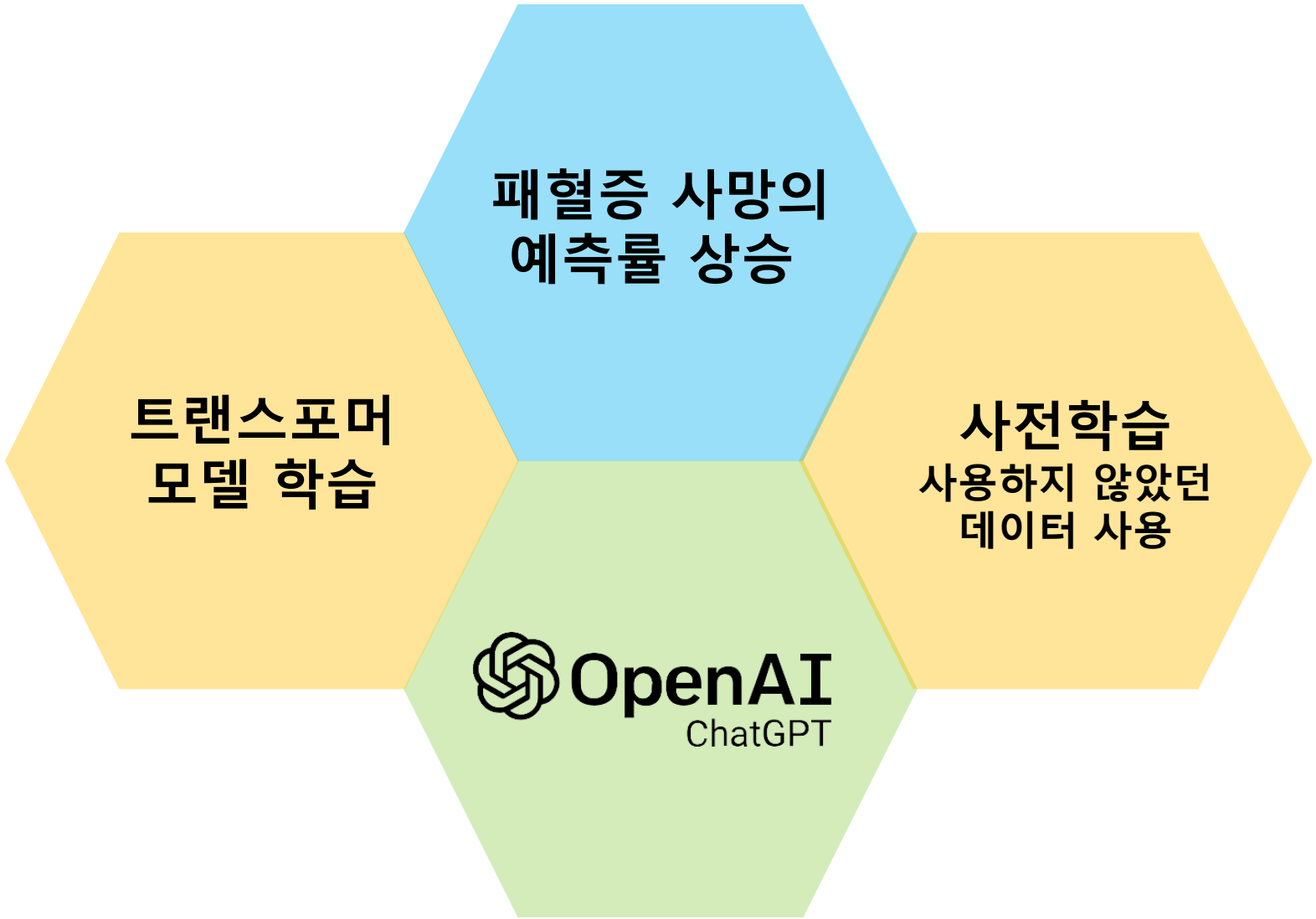
AI트릭스, 패혈증·사망 위험 미리 알려주는 '바이탈케어' 론칭

박정렬 기자 | 입력 2022.12.07 17:48 | 수정 2022.12.07 17:49 | 댓글 0

> 혈압, 맥박 등 생체신호와 혈액 검사 결과 기반
> 전자의무기록 연동돼 의료진 추가 업무 부담 없어
> 기존 방식인 조기경보점수보다 예측 정확도 '우수'
> 비급여 사용 논의, FDA 허가도 진행 계획

식약처의 허가 근거가 된 3건의 임상시험 결과에 따르면 바이탈케어는 일반 병동에서의 급성 중증 이벤트(사망, 중환자실 전실, 심정지), 패혈증, 중환자실에서의 사망 예측 정확도(AUROC)가 각각 0.96, 0.87, 0.98로 기존의 환자 평가 방식인 조기경보점수(NEWS Score)보다 더 높았다.

1.1 분석 동기



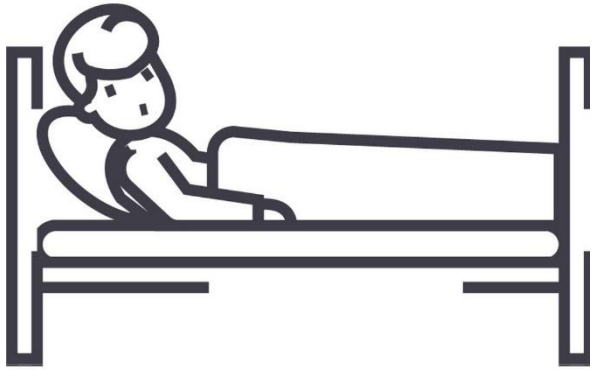
Medical Information Mart for Intensive Care III

- Beth Israel Deaconess Medical Center와 MIT연구자들이 협업하여 만든 데이터
- 2001 ~ 2012년 약 4만명의 비식별 보건의료 데이터

1.2 데이터 설명

ADMISSIONS

ROW_ID	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DEATHTIME	ADMISSION_TYPE	ADMISSION_LOCATION	DISCHARGE_LOCATION	INSURANCE	LANGUAGE	RELIGION	MARITAL_STATUS	ETHNICITY	EDREGTIME
21	22	165315	2196-04-09 12:26:00	2196-04-10 15:54:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	DISC-TRAN CANCER/CHLDRN H	Private	NaN	UNOBTAINABLE	MARRIED	WHITE	2196-04-09 10:06:00
22	23	152223	2153-09-03 07:15:00	2153-09-08 19:10:00	NaN	ELECTIVE	PHYS REFERRAL/NORMAL DELI	HOME HEALTH CARE	Medicare	NaN	CATHOLIC	MARRIED	WHITE	NaN
23	23	124321	2157-10-18 19:34:00	2157-10-25 14:00:00	NaN	EMERGENCY	TRANSFER FROM HOSP/EXTRAM	HOME HEALTH CARE	Medicare	ENGL	CATHOLIC	MARRIED	WHITE	NaN
24	24	161859	2139-06-06 16:14:00	2139-06-09 12:48:00	NaN	EMERGENCY	TRANSFER FROM HOSP/EXTRAM	HOME	Private	NaN	PROTESTANT QUAKER	SINGLE	WHITE	NaN
25	25	129635	2160-11-02 02:06:00	2160-11-05 14:55:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	HOME	Private	NaN	UNOBTAINABLE	MARRIED	WHITE	2160-11-02 01:01:00



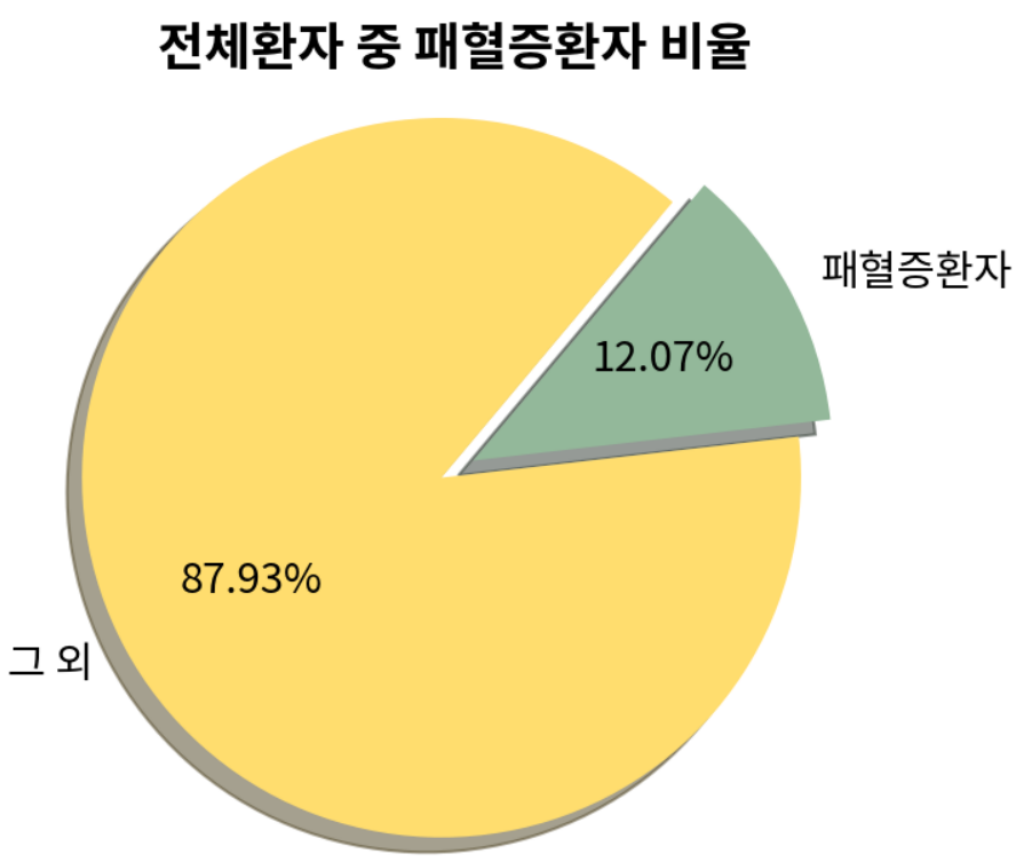
PATIENTS

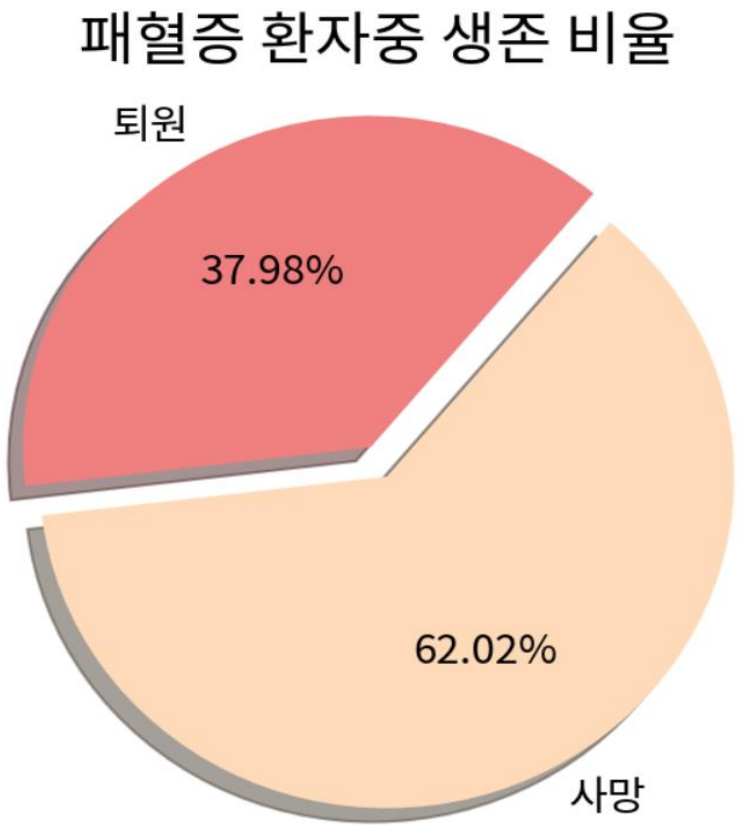
	ROW_ID	SUBJECT_ID	GENDER	DOB	DOD	DOD_HOSP	DOD_SSN	EXPIRE_FLAG
0	234	249	F	2075-03-13 00:00:00	NaN	NaN	NaN	0
1	235	250	F	2164-12-27 00:00:00	2188-11-22 00:00:00	2188-11-22 00:00:00	NaN	1
2	236	251	M	2090-03-15 00:00:00	NaN	NaN	NaN	0
3	237	252	M	2078-03-06 00:00:00	NaN	NaN	NaN	0
4	238	253	F	2089-11-26 00:00:00	NaN	NaN	NaN	0

Part 2.

데이터 전처리

2.1 데이터 전처리 - EDA





2.1 데이터 전처리

Search | Septice(패혈증)

	ROW_ID	ICD9_CODE	SHORT_TITLE	LONG_TITLE
69	242	0031	Salmonella septicemia	Salmonella septicemia
542	593	0545	Herpetic septicemia	Herpetic septicemia
595	646	0380	Streptococcal septicemia	Streptococcal septicemia
598	649	03812	MRSA septicemia	Methicillin resistant Staphylococcus aureus se...
600	651	0382	Pneumococcal septicemia	Pneumococcal septicemia [Streptococcus pneumon...
601	652	0383	Anaerobic septicemia	Septicemia due to anaerobes
602	653	03840	Gram-neg septicemia NOS	Septicemia due to gram-negative organism, unsp...
603	654	03841	H. influenae septicemia	Septicemia due to hemophilus influenzae [H. in...
604	655	03842	E coli septicemia	Septicemia due to escherichia coli [E. coli]
605	656	03843	Pseudomonas septicemia	Septicemia due to pseudomonas
606	657	03844	Serratia septicemia	Septicemia due to serratia
607	658	03849	Gram-neg septicemia NEC	Other septicemia due to gram-negative organisms
608	659	0388	Septicemia NEC	Other specified septicemias
609	660	0389	Septicemia NOS	Unspecified septicemia
653	704	0223	Anthrax septicemia	Anthrax septicemia
6991	7100	65930	Septicemia in labor-unsp	Generalized infection during labor, unspecifie...
9049	9050	77181	NB septicemia [sepsis]	Septicemia [sepsis] of newborn
10304	11403	99591	Sepsis	Sepsis
10305	11404	99592	Severe sepsis	Severe sepsis
13293	13564	67020	Puerperal sepsis-unsp	Puerperal sepsis, unspecified as to episode of...
13294	13565	67022	Puerprl sepsis-del w p/p	Puerperal sepsis, delivered, with mention of p...
13295	13566	67024	Puerperl sepsis-postpart	Puerperal sepsis, postpartum condition or comp...

```
[3]: 99592    3912
     0389     3725
     99591    1272
     03842     467
     03849     395
     0380      376
     77181     225
     0388      206
     03843     127
     03812     118
     0383      110
     0382       88
     03840      60
     03844      27
     03841       4
     0545        2
     0031        1
     Name: ICD9 CODE, dtype: int64
```

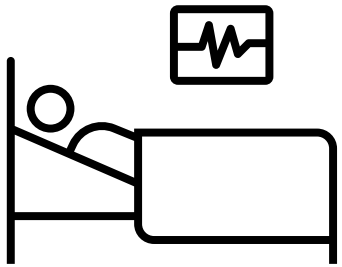
ICD-9 CODE

99592 - Severe sepsis
중증 패혈증

0389 - Unspecified septicemia
불특정 패혈증

99591 - Sepsis
패혈증

2.1 데이터 전처리

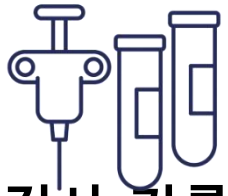


가장 환자가 많은 패혈증 3종 환자
5109명

치료 기록이 없는 환자
3명

5106명

2.1 데이터 전처리



검사 기록

검사종류
검사시각
검사 결과 이상여부

처방 기록

처방 시작시간
처방 종료시간
NDC (의약품 코드)

중환자실 기록

시술 시작시간
시술 종료시간
시술 종류

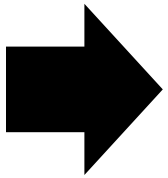
병합된 데이터프레임

정상과 비정상으로 표기

2.2 데이터 병합 - MERGE DATA

total_data

	SUBJECT_ID	ITEMID	CHARTTIME	FLAG	TYPE
0	3	50912	2101-10-04	abnormal	LAB
1	3	50931	2101-10-04	abnormal	LAB
2	3	51006	2101-10-04	abnormal	LAB
3	3	51221	2101-10-04	abnormal	LAB
4	3	51222	2101-10-04	abnormal	LAB
...	...	...	...	...	...
7679185	99991	904150061	2185-01-05	NaN	PRE
7679186	99991	54839224	2185-01-05	NaN	PRE
7679187	99991	456066270	2185-01-05	NaN	PRE
7679188	99991	58177020211	2185-01-05	NaN	PRE
7679189	99991	63481062375	2185-01-05	NaN	PRE



	SUBJECT_ID	CHARTTIME	50803	50804	50805	50806	50808	50809	50811	50813	50814
1	3	2101-10-04	0	0	0	0	0	0	0	0	0
2	3	2101-10-05	0	0	0	0	0	0	0	0	0
3	3	2101-10-06	0	0	0	0	0	0	0	0	0
4	3	2101-10-07	0	0	0	0	0	0	0	0	0
5	3	2101-10-11	0	0	0	0	0	0	0	0	0
6	3	2101-10-12	0	0	0	0	1	1	1	0	0
7	3	2101-10-13	0	0	0	0	0	0	0	0	0
8	3	2101-10-14	0	0	0	0	0	0	0	0	0
9	3	2101-10-15	0	0	0	0	0	0	0	0	0
10	3	2101-10-16	0	0	0	0	0	0	0	0	0
11	3	2101-10-18	0	0	0	0	0	0	0	0	0
12	3	2101-10-20	0	0	0	1	1	1	1	1	0
13	3	2101-10-21	1	0	0	0	1	1	1	1	0
14	3	2101-10-22	0	0	0	0	0	0	1	0	0
15	3	2101-10-23	0	0	0	0	1	0	0	0	0
16	3	2101-10-24	0	0	0	0	0	0	1	0	0
17	3	2101-10-25	0	0	0	0	0	0	0	0	0
18	3	2101-10-26	0	0	0	0	0	0	0	0	0
19	3	2101-10-27	0	0	0	0	0	0	0	0	0
20	3	2101-10-28	0	0	0	0	0	0	0	0	0
21	3	2101-10-29	0	0	0	0	0	0	0	0	0

- 환자 x 날짜 행과 ITEM_ID 열로 구성된 Zero matrix에 값 채워넣기
- data.csv로 사전학습 데이터(전체 데이터) 저장

Part 3.

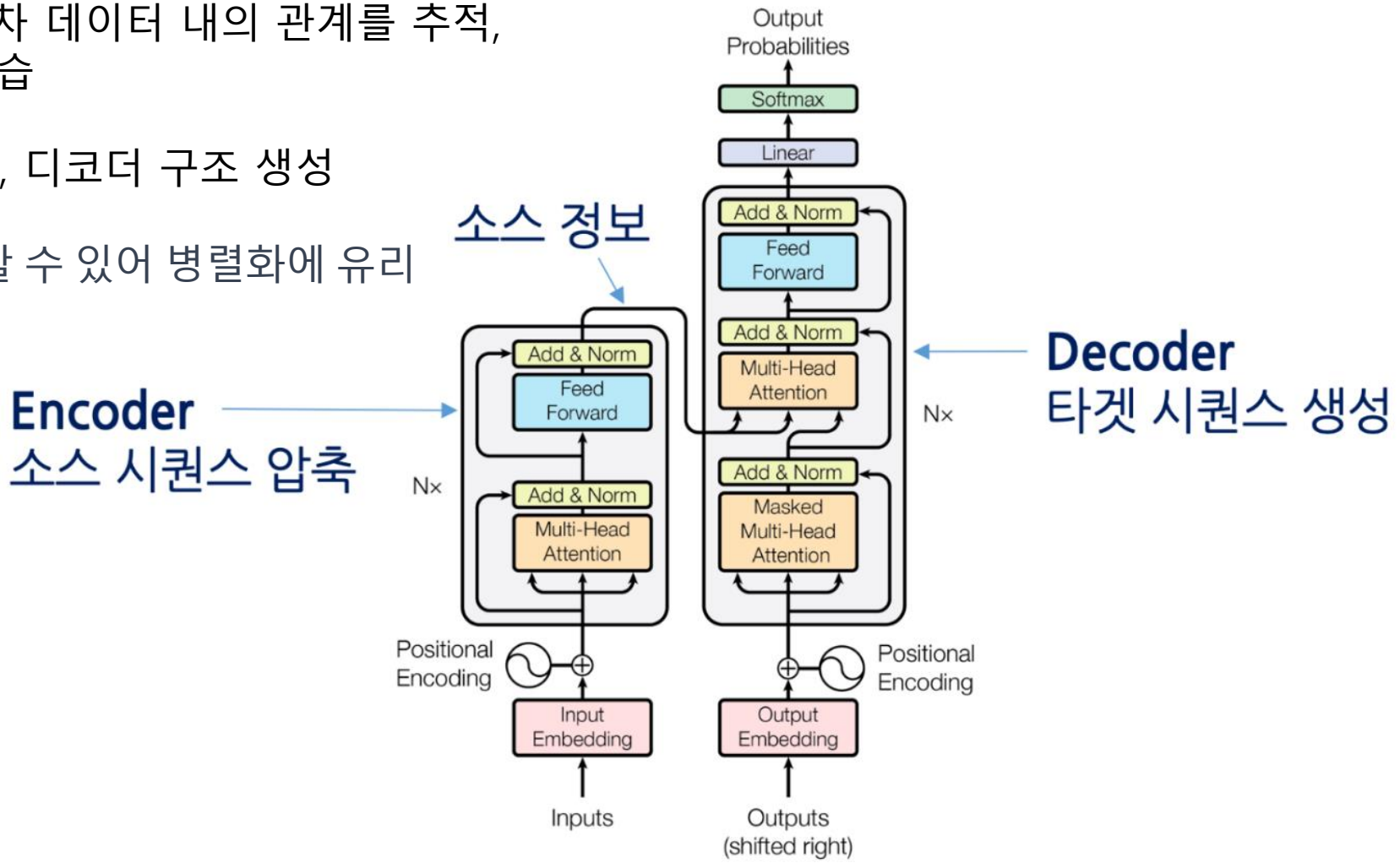
사전학습, 전이학습

트랜스포머

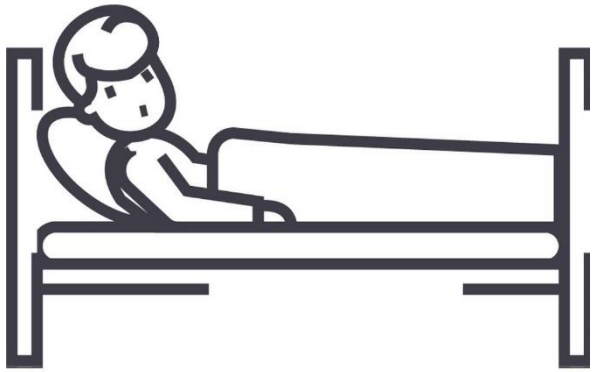
Attention-mechnism 을 활용해 순차 데이터 내의 관계를 추적,
시계열 데이터의 복잡한 패턴을 학습

Attention-mechnism 만으로 인코더, 디코더 구조 생성

시퀀스의 모든 부분을 동시에 처리할 수 있어 병렬화에 유리

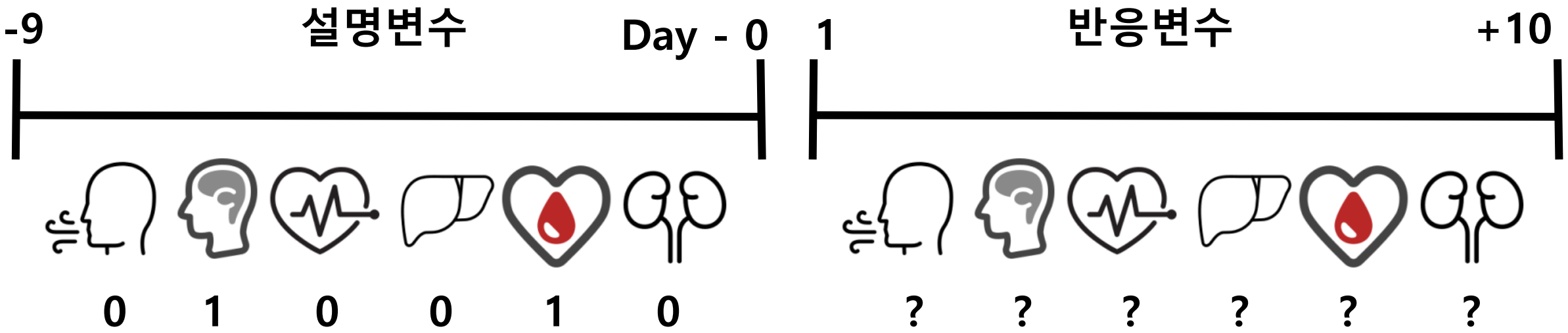


3.2 사전학습 데이터

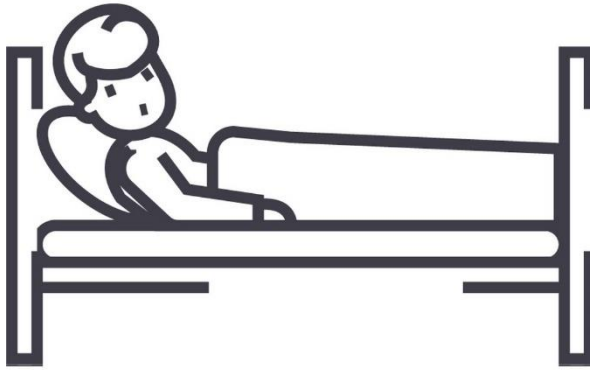


ex) Subject ID - 3

기준일 : 2123 - 05 - 21

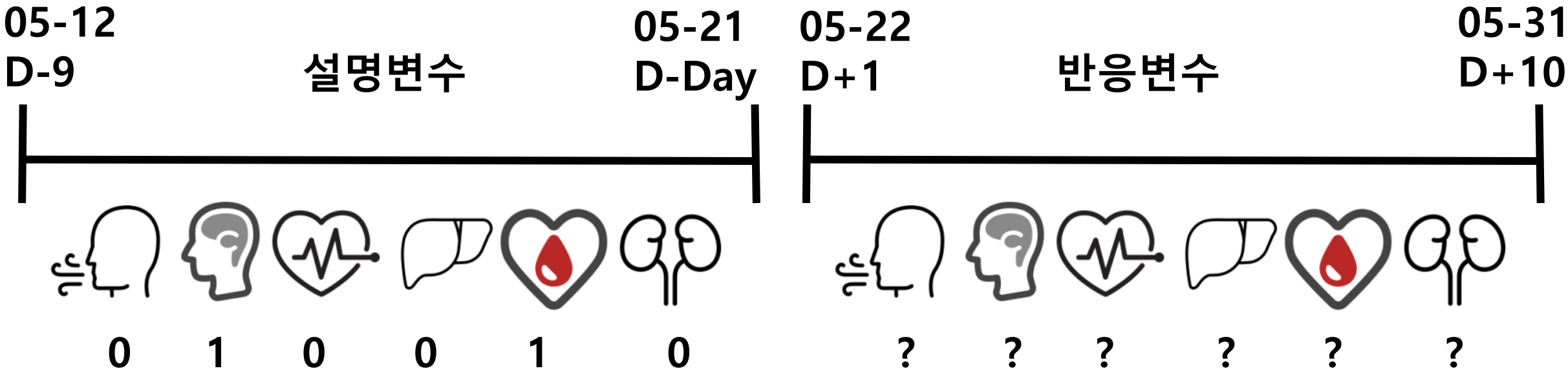


3.2 사전학습 데이터



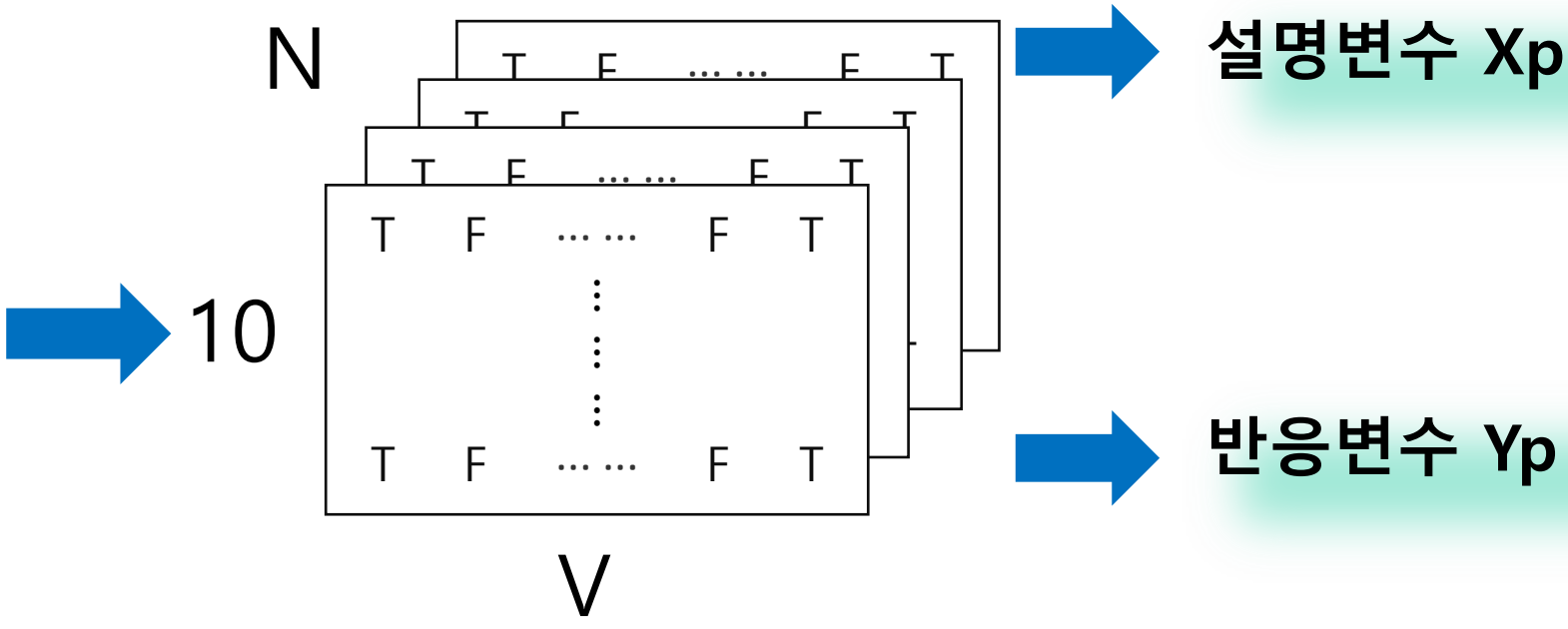
ex) Subject ID - 3

기준일 : 2123 - 05 - 21



학습효율을 위해 0 / 1 을 True / False 로 변경

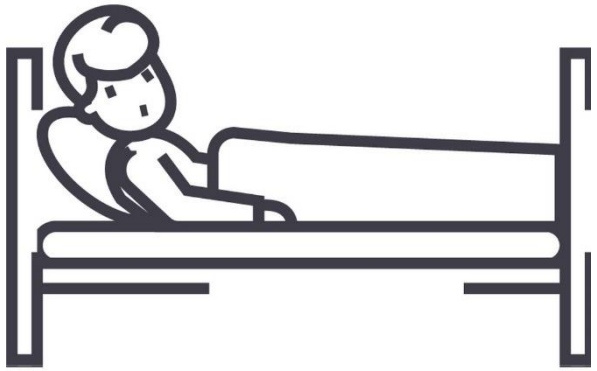
	SUBJECT_ID	CHARTTIME	50803	50804	50805	50806	50808
1	3	2101-10-04	0	0	0	0	0
2	3	2101-10-05	0	0	0	0	0
3	3	2101-10-06	0	0	0	0	0
4	3	2101-10-07	0	0	0	0	0
5	3	2101-10-11	0	0	0	0	0
6	3	2101-10-12	0	0	0	0	1
7	3	2101-10-13	0	0	0	0	0
8	3	2101-10-14	0	0	0	0	0
9	3	2101-10-15	0	0	0	0	0
10	3	2101-10-16	0	0	0	0	0
11	3	2101-10-18	0	0	0	0	0
12	3	2101-10-20	0	0	0	1	1
13	3	2101-10-21	1	0	0	0	1
14	3	2101-10-22	0	0	0	0	0
15	3	2101-10-23	0	0	0	0	1
16	3	2101-10-24	0	0	0	0	0
17	3	2101-10-25	0	0	0	0	0
18	3	2101-10-26	0	0	0	0	0
19	3	2101-10-27	0	0	0	0	0
20	3	2101-10-28	0	0	0	0	0
21	3	2101-10-29	0	0	0	0	0



데이터개수N * 10 * 변수 수 V

Data 총 개수 112937 x 10 x 283의 3차텐서

3.3 전이학습 데이터



Subject ID - 3

05-11
D-10

설명변수

05-20
D-1

0 1 0 0 1 0

2123 - 05 - 21
D - Day

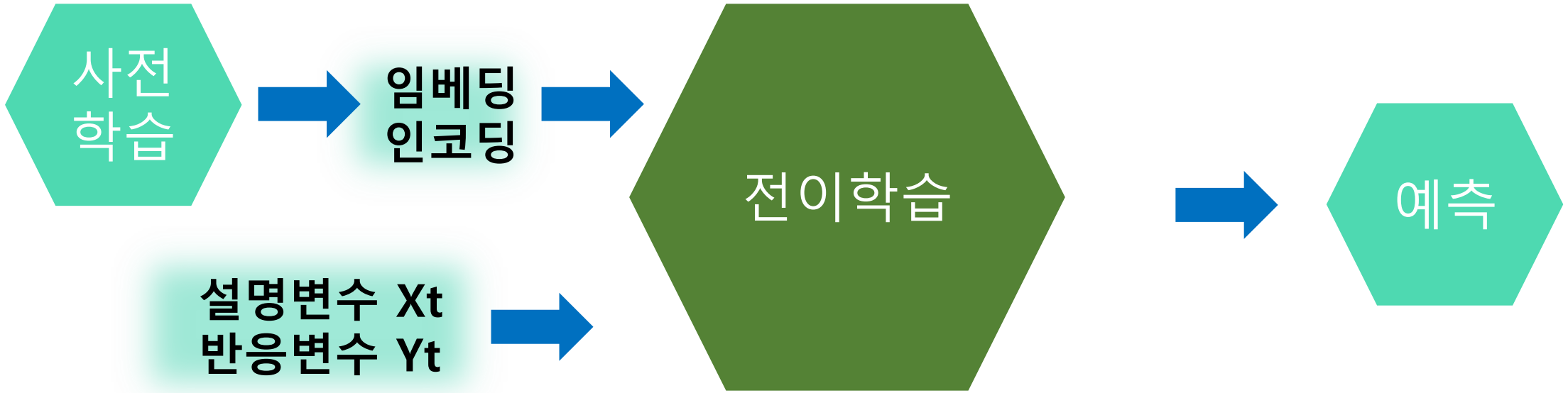
퇴원/사망일

반응변수

0 퇴원

1 사망

3.3 전이학습 모델



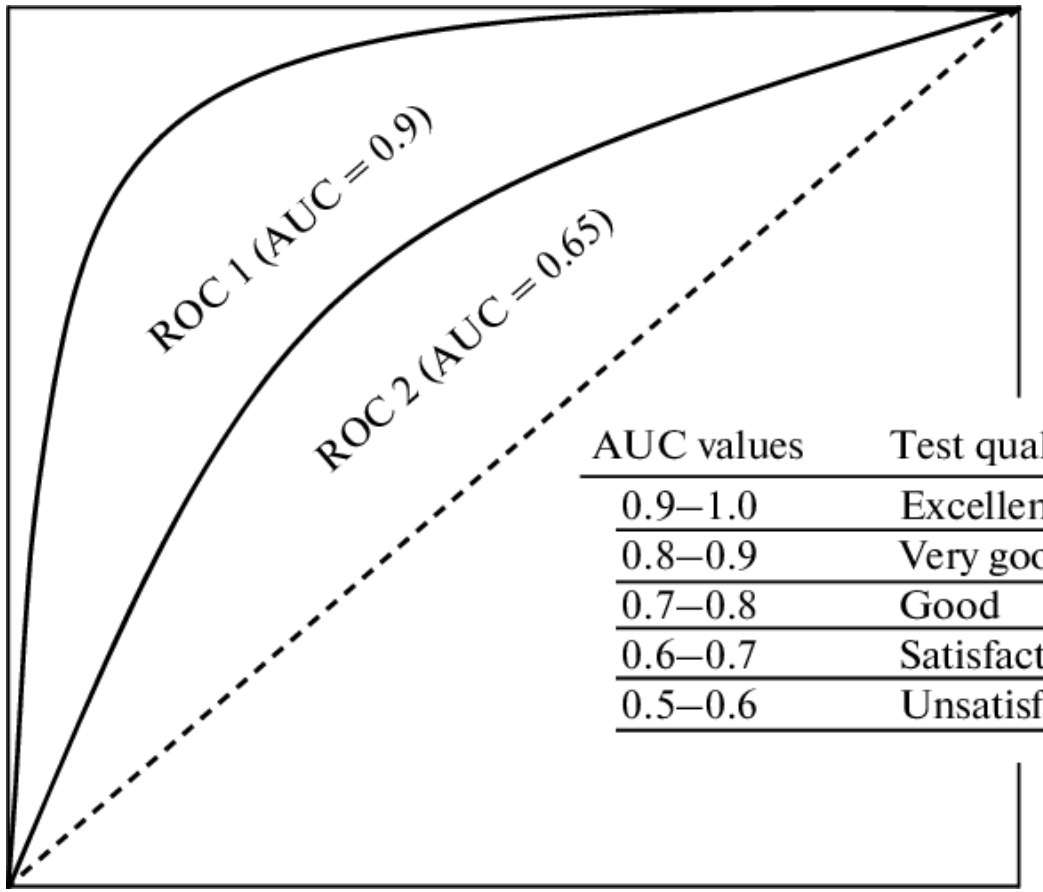
ROC - AUC

모든 가능한 분류 임계값에 대해 TPR(True Positive Rate)과 FPR(False Positive Rate)을 기반으로 모델의 성능을 다양한 임계값에서 평가

클래스 불균형에 영향을 덜 받음

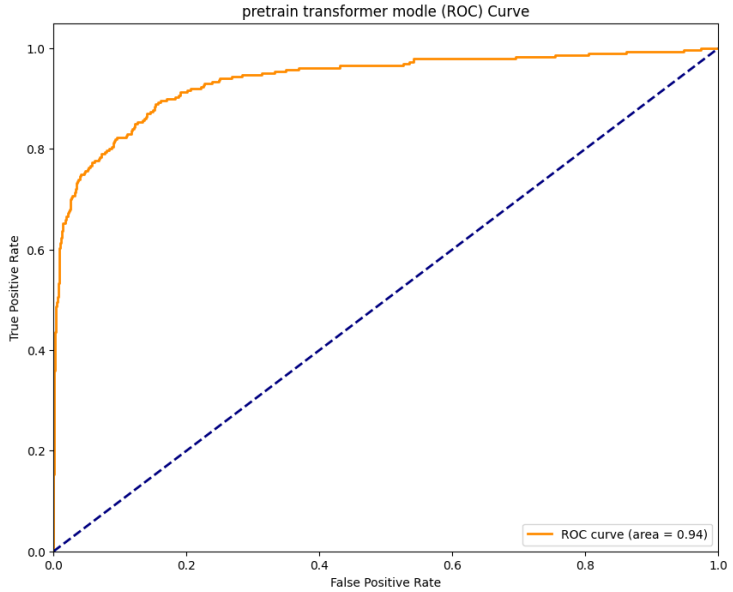
모델의 성능을 보다 객관적 평가 가능

데이터의 사망, 생존율 불균형



AUC values	Test quality
0.9–1.0	Excellent
0.8–0.9	Very good
0.7–0.8	Good
0.6–0.7	Satisfactory
0.5–0.6	Unsatisfactory

3.4 검증

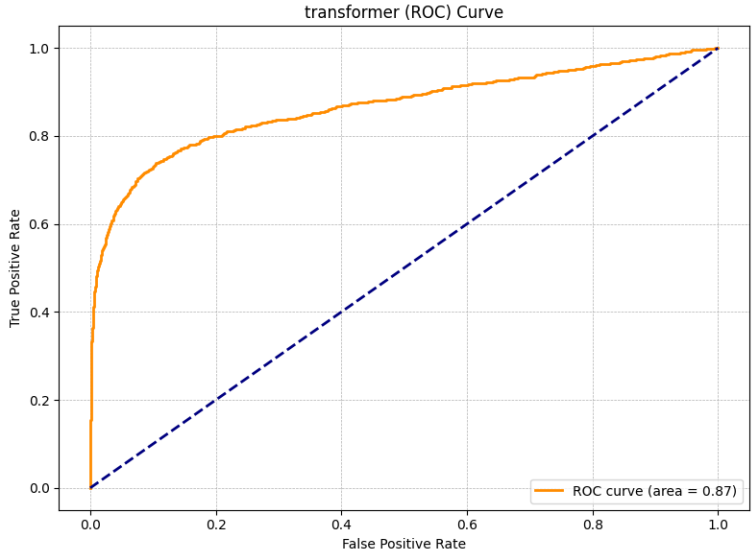


전이학습 트랜스포머 모델 학습

환자의 내일 생존/ 사망 예측

ROC – AUC

94%



트랜스포머 모델 학습

환자의 내일 생존/ 사망 예측

ROC – AUC

87%

환자의 내일 생존/ 사망 예측에 있어
ROC – AUC : 93.47% 의 정확도를 보여준다.

- 사전학습을 하지 않은 트랜스포머 모델의 경우 88%
- 기존의 모델보다 월등히 높은 정확도를 보여준다.

Part 4.

결론

4.1 한계 및 기대효과

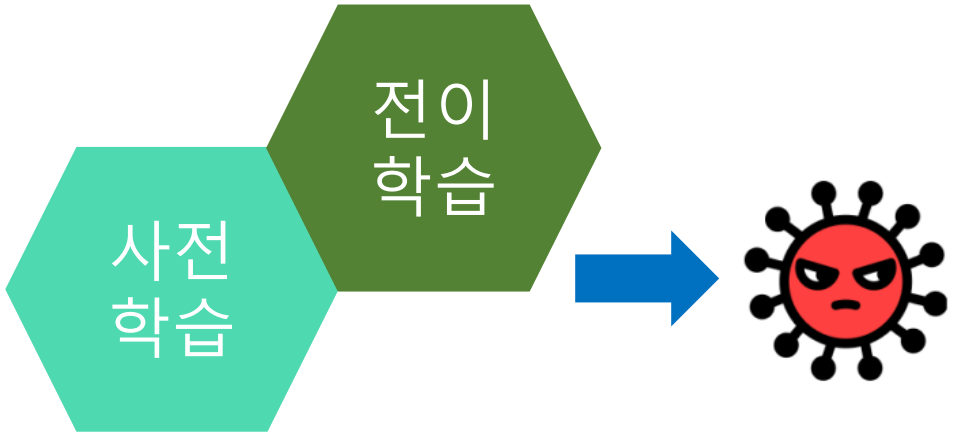
모델의 한계

- 수백만 개의 데이터를 학습시키는 과정에서, Bool 형으로 데이터를 변환하거나 일부 열을 삭제하는 등 데이터 손실이 발생했다. 이로 인해 데이터와 모델의 효율을 한계까지 활용하지 못했다.
- 중환자실 데이터를 이용해 총 283개의 검사수치를 학습에 활용했다. 검사의 종류와 가짓수가 다른 일반 병동 등의 환경에서의 사용은 제한될 수 있다.

4.1 한계 및 기대효과

환자의 패혈증 위험도에 대해
빠르게 알려 즉각 조치 가능

사전학습 - 전이학습 모델을
다른 질환에 적용하여 사용 가능하다.



4.2 출처

당뇨증가 : 의협신문 https://www.doctorsnews.co.kr/news/articleView.html?idxno=146606

고령증가 : 통계청 https://kiri.or.kr/PDF/weeklytrend/20221011/trend20221011_4.pdf

패혈증시장증가 : 링트인 https://kr.linkedin.com/pulse/sepsis-diagnostics-market-tushar-jane

감사합니다

Thank You