Critical event prediction at NICU

SNUH 분당서울대학교병원
SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

1. Scientific overview

Challenge objective

SNUH 분당서울대학교병원
SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

연구 대상

- 분당서울대학교병원 신생아 중환자실 (NICU)에 입원한 저체중 미숙아
- 출생 체중 1,500 g 미만
- 재태주수 32주 미만

Challenge objective

연구 배경 – 미숙아 비율의 증가

- 산모의 고령화, 환경 유해 인자 증가 등으로 인해 미숙아 출생 비중이 증가
- 1993-2005년 미국의 저체중 미숙아의 비율이 점차적으로 증가

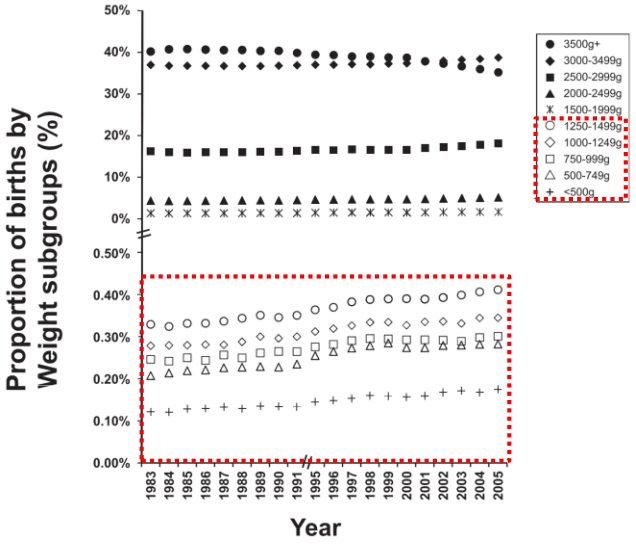


FIGURE 1
Percentage contributions of infants in each VLBW subgroup to total births for 1983–2005. The contribution of infants in each VLBW subgroup increased during the study period.

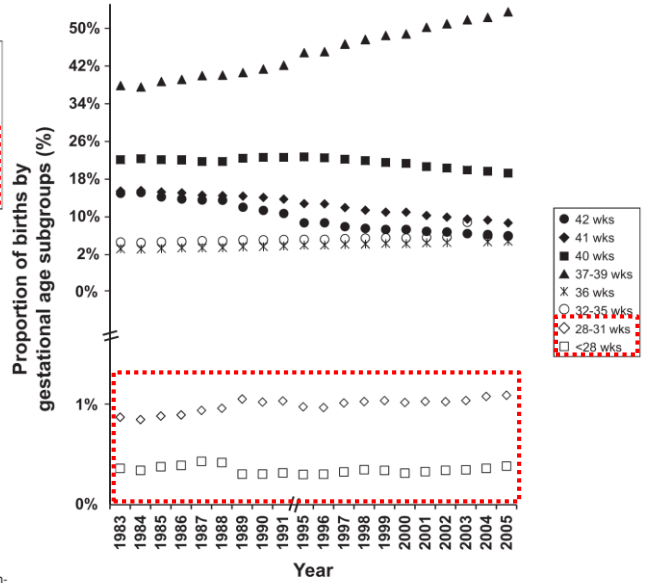


FIGURE 2
Percentage contributions of infants in gestational age subgroups to total births for 1983–2005.

Lau C, Ambalavanan N, Chakraborty H, Wingate MS, Carlo WA. Extremely low birth weight and infant mortality rates in the United States. *Pediatrics*. 2013;131(5):855-860. doi:10.1542/peds.2012-2471

Challenge objective

연구 배경 – 미숙아 발병 및 사망률

- 치료기술의 발달로 전체 신생아의 생존율 크게 증가
- 그러나 극소저체중출생아의 경우 여러 합병증과 감염 발생 위험이 높음
- 1993-2005 년도의 재태일수 28주 미만의 미숙아의 사망 비중이 증가

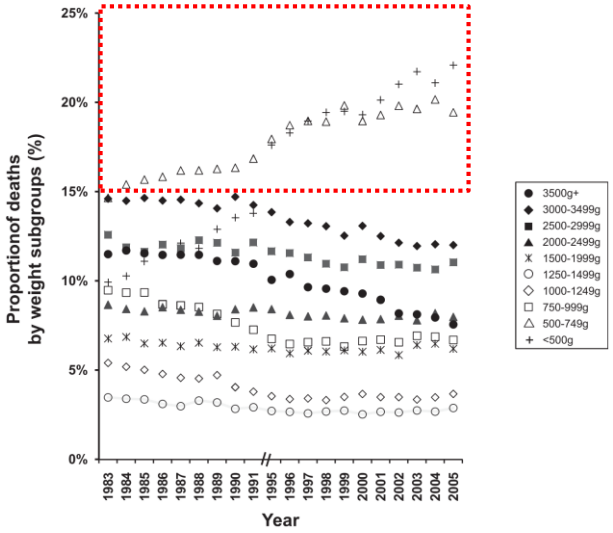


FIGURE 3
Percentage contribution of infants in each birth weight subgroup to infant mortality for 1983–2005. The subgroups that contributed the most to infant mortality were the VLBW subgroups despite their relatively small contribution to births. Infants <500 g contributed the most (>20%) to deaths in recent years.

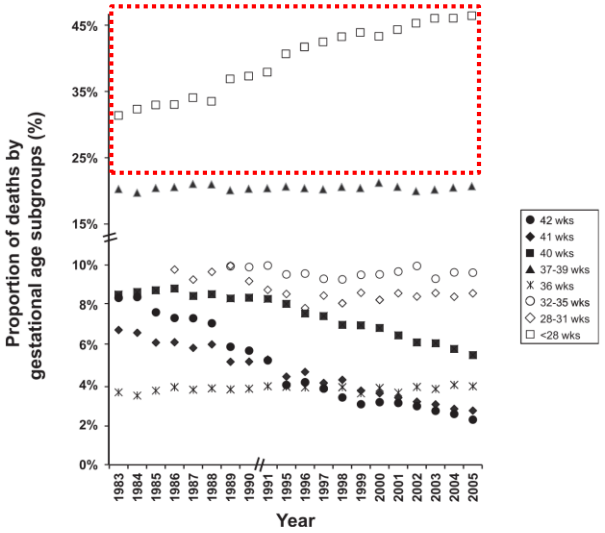


FIGURE 4
Percentage contribution of infants in gestational age subgroups to infant mortality for 1983–2005. Infants, 28 weeks contributed the most to infant mortality, and their contribution increased over the years.

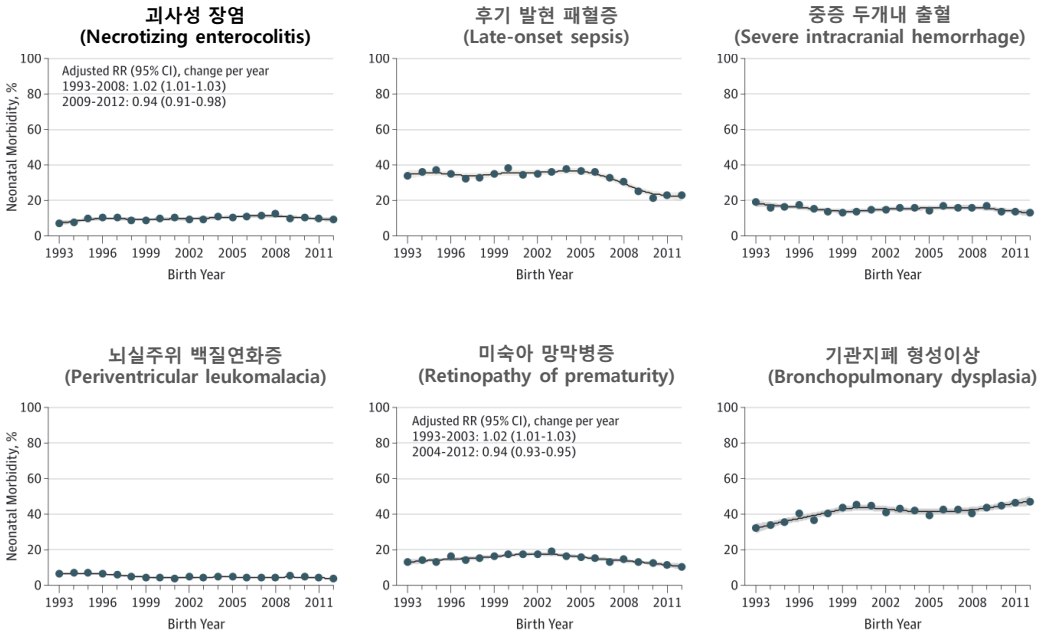
Lau C, Ambalavanan N, Chakraborty H, Wingate MS, Carlo WA. Extremely low birth weight and infant mortality rates in the United States. *Pediatrics*. 2013;131(5):855-860. doi:10.1542/peds.2012-2471

Challenge objective

연구 배경 – 미숙아 발병 및 사망률

- 재태 일수 28주 이하 미숙아 내에서도 1993-2012 년까지의 연도별 RR이 유의미하게 변화하지 않음
- 주요 발병 요인 없이 생존하여 퇴원한 저숙아의 비중이 크게 변화하지 않음
- 괴사성 장염(Necrotizing enterocolitis), 후기 발현 패혈증(Late-onset sepsis), 중증 두개내 출혈(Severe intracranial hemorrhage), 뇌실주위 백질연화증 (Periventricular leukomalacia), 미숙아 망막병증(Retinopathy of prematurity), 기관지폐 형성이상(Bronchopulmonary dysplasia) 비율 변화가 크지 않음

Figure 2. Neonatal Morbidities for Infants Born at Gestational Ages 22 Through 28 Weeks



Stoll BJ, Hansen NI, Bell EF, et al. Trends in care practices, morbidity, and mortality of extremely preterm Neonates, 1993-2012. *JAMA - J Am Med Assoc.* 2015;314(10):1039-1051. doi:10.1001/jama.2015.10244

Challenge objective

SNUH 분당서울대학교병원 SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

연구 배경 – 한국내 미숙아 발병 및 사망률

- 한국의 경우 미국의 산모의 나이에 비해 높으며 미숙아와 저체중 출생아의 비율이 낮음
- 한국의 전체 신생아 사망율이 미국보다 낮으나 미숙아와 저체준 출생군에서는 한국이 미국보다 높은 것으로 나타남, 이는 미국에 비해 저체중 미숙아의 수가 차지하는 비율이 낮기 때문에 나타남

Table 1. Characteristics of Live Births in South Korea (2009–2011) and Those in the United States (2000–2011)

	Korea 2009-2011	United States, 2000-2010				*P*-value
		Korean	White	Black	Total	
Live births (n)	1,383,806	107,309	31,588,183	4,381,664	49,384,187	0.00
Birth weight (mean, g)	3,216	3,270	3,341	3,116	3,281	0.00
Gestational age (mean, wk)	38.71	39.0	38.7	38.2	38.6	0.00
<37 weeks (%)	6.0	6.4	10.8	16.3	12.1	0.00
<2,500 g (%)	5.1	4.9	6.6	12.4	7.9	0.00
Maternal age (mean, yr)	30.7	30.9	27.1	25.6	26.4	0.00
>12 yr of education (%)	67.0	86.4	54.4	46.0	48.9	0.00
Multiple births (%)	2.8	2.5	3.5	3.7	3.3	0.00
Married (%)	97.8	97.6	76.7	44.1	62.6	0.00

Table 3. Birth Weight Distribution and Birth Weight-Specific Neonatal Mortality Rate, and Gestational Age Distribution and Gestational Age-Specific Neonatal Mortality Rate

	Korea 2009–2011	United States, 2000–2011							
		Korean		White		Black		Total	
BW (kg)	BWD, %	NMR	BWD, %	NMR	BWD, %	NMR	BWD, %	NMR	BWD, %	NMR
<1.0	0.3	275.8	0.3	228.7	0.4	293.6	1.4	256.4	0.6	293.4
1.0–1.49	0.4	60.4	0.4	34.5	0.6	40.8	1.3	30.5	0.7	39.4
1.5–1.99	0.8	15.1	0.9	7.7	1.3	17.6	2.4	12.3	1.6	17.0
2.0–2.49	3.7	3.8	3.4	3.9	4.2	5.9	7.3	4.2	5.0	5.6
>2.5	94.9	0.5	95.1	0.4	93.4	0.7	87.6	0.9	92.0	0.8
GA (wk)	GAD, %	NMR	GAD, %	NMR	GAD, %	NMR	GAD, %	NMR	GAD, %	NMR
<28	0.3	271.6	0.3	232.2	0.4	278.2	1.3	242.5	0.6	272.3
28–31	0.5	48.3	0.6	26.4	1.0	31.0	2.0	28.4	1.2	30.8
32–36	5.2	5.1	5.6	3.5	9.4	4.7	13.0	4.3	10.2	4.8
≥37	94.1	0.5	92.6	0.4	88.6	0.8	83.3	1.0	87.3	0.9
All	100.0	1.7	100.0	1.3	100.0	2.7	100.0	5.3	100.0	3.4

Abbreviations: BW, birth weight; BWD, birth weight distribution; NMR, neonate mortality rate; GA, gestational age; GAD, gestational age

Lee JS, Yoon JM, Cheon EJ, et al. Comparison of Neonatal and Infant Mortality Rates between Infants Born in South Korea and Those Born in the United States. *Neonatal Med.* 2014;21(4):211. doi:10.5385/nm.2014.21.4.211

Challenge objective

연구 배경 – 생체 신호 정상범위

- 신생아의 생체 신호는 평균적으로 출생 후 3일에서 6일사이의 정상 범위를 구할 수 있으나 그 이후의 생체 신호는 발산하여 표준 정상 생체신호 범위를 구성하기 어려움
- 미숙아의 경우 발산 정도가 더 크며, 환자의 상태를 정확하게 진단하기 위해서는 환자의 지속적인 관찰과 기저 질환, 의료 중재에 따른 영향성에 대해 높은 이해도를 요구

Fig. 5 Systolic, mean and diastolic blood pressures relative to birth weight on day 1 of life [the *boxes* show 5th and 95th confidence intervals (CI), the *horizontal bar* indicates the median and the *longitudinal lines* represent the range of values]. *Blue* = diastolic, *red* = mean, *green* = systolic

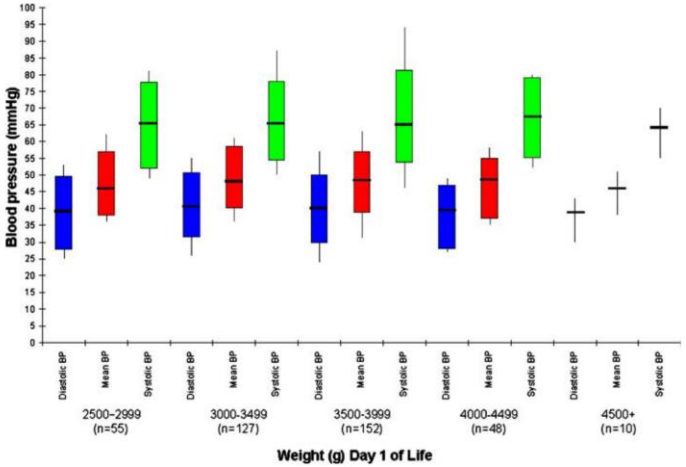
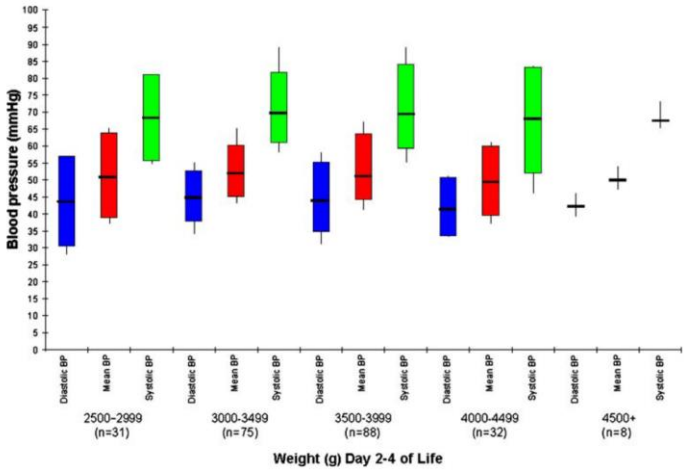


Fig. 6 Systolic, mean and diastolic blood pressures relative to birth weight on days 2–4 of life [the *boxes* show 5th and 95th confidence intervals (CI), the *horizontal bar* indicates the median and the *longitudinal lines* represent the range of values]. *Blue* = diastolic, *red* = mean, *green* = systolic



Challenge objective

연구 배경 – 기존 예측 모델

- 이미 인공지능을 활용하여 질병 예방 및 예측하는 연구가 다양하게 진행
- 존스 홉킨스 의과대학 연구진은 중환자실 환자 대상으로 매일 기록되는 체온, 호흡수, 심장박동수, 혈압 등의 일상 데이터를 바탕으로 Septic shock를 예측하는 실시간 조기 경고 점수 개발
- 신생아의 경우 heart rate characteristics (HRC)의 분석 결과를 활용하여 HeRO (heart rate observation) monitoring system의 임상 시험을 수행하여 중증 질환 예측이 가능함을 보여주었으나, 수술, pancuronium (paralytic agents), atropine (Anticholinergic agents), Dexamethasone 등의 투여로 heRO score 가 달라지는 한계점이 있으며 연구 결과에서도 호흡 패턴, 기타 생체 신호로 보완해야 할 것을 기술

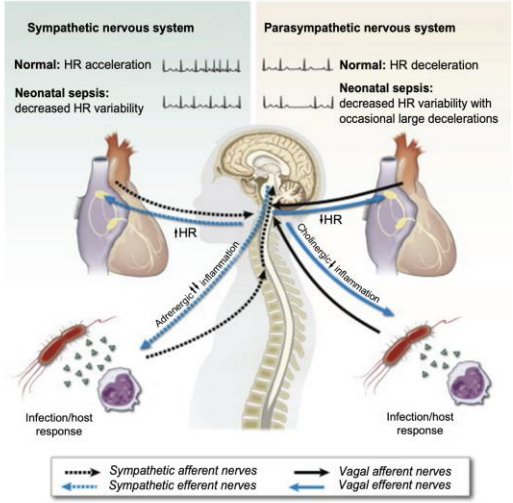


Figure 1 Effects of autonomic nervous system activation on heart rate characteristics and host response in sepsis.
Notes: In the healthy state, sympathetic activation leads to release of norepinephrine, which increases heart rate, and parasympathetic (vagus nerve) activation leads to release of acetylcholine, which decreases heart rate. Pathogens and inflammatory cytokines can activate the autonomic nervous system, which can in turn alter the inflammatory cytokine cascade. Sepsis can lead to altered sensitivity to adrenergic and cholinergic stimuli (fewer heart rate accelerations and exaggerated decelerations).

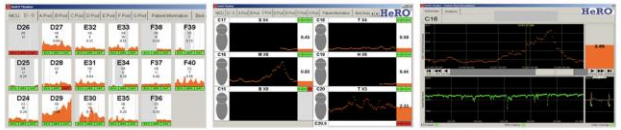


Figure 2 HeRO monitor screens.
Notes: Three displays of the heart rate characteristics index (HeRO score). Left: unit view shows 5-day HeRO score trends (orange) for multiple patients. Center: pod view shows 5-day HeRO score trends (orange line) and current HeRO score in patients from a particular area of the neonatal intensive care unit. Right: individual patient view (see also Figure 3).
Abbreviation: HeRO, heart rate observation.

Fairchild K, Aschner. HeRO monitoring to reduce mortality in NICU patients. *Res Reports Neonatol.* 2012:65. doi:10.2147/rrn.s32570

Challenge objective

SNUH 분당서울대학교병원 SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

연구 목표 – Critical Events

- 본 예측모델에서는 저혈압, 패혈증, 괴사성 장염, 사망을 critical event 로 정의
- 저혈압 (Hypotension)
    - 평균 혈압이 30 mmHg 미만, 이완기 혈압 20 mmHg 미만 대상
    - 승압제 또는 hydrocortisone 오더가 접수된 시점을 이벤트 시점으로 정의
- 패혈증 (Sepsis)
    - Blood culture 에서 일반 세균이 검출되고 해당 시점 2일 이내 항생제가 5일 이상 투여되는 경우
    - 최초로 일반 세균이 검출된 blood culture 검사 채혈 일시를 이벤트 시점으로 정의
- 괴사성 장염 (Necrotizing enterocolitics)
    - 입원 기간 중 괴사성 장염으로 진단된 환아
    - Infantogram 및 abdomen sonograph 에서 괴사성 장염에 관련된 판독결과가 기록된 경우
    - 금식 7일 이상 시행된 경우
    - 항생제 처방이 7일 이상 투여된 경우
    - Infantogram, abdomen sonograph에서 괴사성 장염이 최초로 확인된 시점에서 금식 시작 시간 또는 항생제 오더 시점을 이벤트 시점으로 정의
- 사망 (Mortality)
    - 신생아 입원 중 원내 사망 시점을 이벤트 시점으로 정의

Challenge objective

SNUH 분당서울대학교병원 SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

연구 목표 – 예측 모델 목표

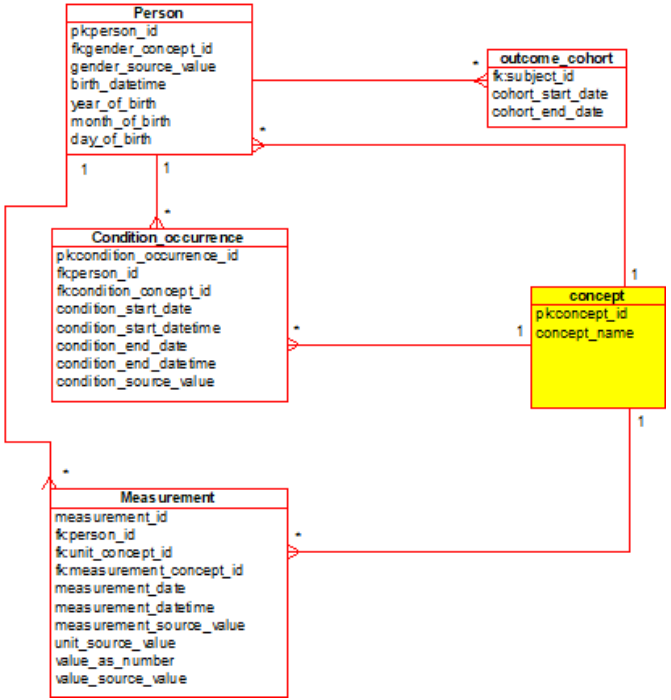
- 본원의 신생아 중환자실에서는 환자의 상태를 모니터링하기 위한 다양한 생체신호 데이터 측정 및 수집
- 최소 30초 단위로 기록되는 환자 생체 신호 데이터를 분석, 판단하여 12시간 이후의 critical events를 예측하는 모델을 개발
- 예측의 정확도 및 Positive predictive value 동시에 확보, 현장 도입 시 실제 신생아의 조기 중재 및 false alarm 피로도를 낮추어 의료의 질을 높이는 것이 주 목적

2. Challenge Data

Challenge Data

Data Format & Characteristics

- 데이터는 환자의 실시간 생체 신호 정보, 진단 정보, 예측을 위한 코호트가 기록된 데이터베이스의 csv 파일로 구성
- 테이블 구성
    - CONDITION_OCCURRENCE
    - MEASUREMENT
    - OUTCOME_COHORT
    - PERSON
    - CONCEPT



Challenge Data

PERSON Table

- 환자의 demographics 및 고유 식별자 (person_id) 기술
- demographics
    - 성별
    - 출생년도, 출생월일
    - 국적, 인종
- 고유 식별자인 person_id 로 measurement, condition_occurrence 테이블에서 특정 환자 정보 조회 가능

Field	Type	Description
person_id	integer	각 신생아 별 고유 식별자
birth_datetime	datetime	신생아 출생 생년일시 (2019-10-11 22:10:30)
gender_source_value	string	신생아 성별 (‘M’, ‘F’)
year_of_birth	integer	출생 연도 (2019)
month_of_birth	integer	출생 월 (10)
day_of_month	integer	출생 일자 (11)
gender_concept_id	integer	신생아 성별을 기술하는 concept table 의 foreign key

Challenge Data

CONDITION_OCCURRENCE Table

- 환자의 medical condition, diseases, diagnosis, sign, symptom 이 기록
- 'condition_source_value' 는 KCD로 작성
- 'condition_concept_id' 의 ICD-10 code 조회하여 확인 가능
- 예측 내용에 직접적인 연관성을 지닌 진단 코드는 제외
    - P36.9 (Bacterial sepsis of newborn, unspecified)
    - P77 (Necrotizing enterocolitis of fetus and newborn)
    - R57.2 (Septic shock) ...

Field	Type	Description
condition_occurrence_id	integer	condtion_occurrence 고유 식별자
person_id	integer	대상 신생아의 perosn_id foreign key
condition_start_date	date	진단 등록 시점 날짜 (2019-10-30)
condition_start_datetime	datetime	진단 등록 시점 날짜 및 시간 (2019-10-30 11:30:30)
condition_source_value	string	KCD 진단 코드(P22.0)
condition_concept_id	int	KCD와 대응되는 ICD10 concept_id foreign key (45597070)

Challenge Data

MEASUREMENT Table

- Patient monitor 에서 측정 및 기입된 생체 신호 정보 기록
- 측정 항목
  - measurement_source_value
  - measurement_concept_id
- 측정 값
  - value_as_number : float type
  - value_source_value : raw string data
- 단위
  - unit_concept_id
  - unit_source_value : raw string data

Field	Type	Description
measurement_id	integer	measurement 고유 식별자
person_id	integer	대상 신생아의 perosn_id foreign key
measurement_concept_id	integer	측정 항목에 대응하는 concept_id
measurement_date	date	생체 신호 측정 날짜 (2019-10-30)
measurement_datetime	datetime	생체 신호 측정 날짜 및 시간 (2019-11-11 11:30:30)
value_as_number	float	생체 신호 수치형 측정값 (96.0)
measurement_source_value	string	측정된 생체 신호 항목 이름 (SpO2)
measurement_concept_id	int	측정 생체 신호 concept_id foreign key
unit_source_value	string	측정 단위 (%)
value_source_value	string	원본 측정 값 문자열 (‘96.0’)

Challenge Data

SNUH 분당서울대학교병원 SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

MEASUREMENT Table

- 생체 신호는 다양한 기기에서 수집되므로 동일한 항목에 대해 여러 이름으로 표기

항목 이름	measurement_source_value 표기	설명
IDBP	ARTd, ABPd	Invasive diastolic arterial pressure
IMBP	ABPm	Invasive mean arterial pressure
ISBP	ARTs, ABPs	Invasive systolic arterial pressure
FDBP	AoD	Femoral artery Diastolic blood pressure
FMBP	AoM	Femoral artery Mean blood pressure
FSBP	AoS	Femoral artery Systolic blood pressure
BT	Tskin, Trect, Tnaso, Tesoph, Temp, Tcore	Body Temperature
CVP	CVPm	Central Venuous pressure
ETCO2	etCO2	End tidal carbon dioxide concentration
PR	HR, Pulse	Heart rate
LAP	LAPm	Mean direct left atrial pressure
MINUTE_VOLUME	MINVOL, MV	Airway Minute Volume Inspiratory
PMEAN	MnAwP, Pmean	Mean inspiratory airway pressure
DBP	NBPd, NBP-D	Non-invasive diastolic arterial pressure
MBP	NBPm, NBP-M	Non-invasive mean arterial pressure
SBP	NBPs, NBP-S	Non-invasive systolic arterial pressure
DPAP	PAPd, Pd	Pulmonary artery Diastolic blood pressure
MPAP	PAPm, Pm	Pulmonary artery Mean blood pressure
SPAP	PAPs, Ps	Pulmonary artery Systolic blood pressure
PPEAK	PIP, Ppeak	Peak inspiratory pressure
RR	RR, Resp	Respiratory rate
FREQ_MEASURE	RRaw	Airway respiration Rate
SPO2	SpO2T,SpO2-%,SpO2	Hemoglobin saturation with oxygen
VTE	TV	Tidal volume
VIT	TVin	Inspiratory tidal volume

Challenge Data

MEASUREMENT Table

- 생체 신호는 다양한 기기에서 수집되므로 동일한 항목에 대해 여러 이름으로 표기
- 측정 단위 또한 측정 기기별 정의가 다르므로 동일 단위에 대한 표기가 다름
- 본 데이터에서는 한 측정 항목에 대해 단위 차이가 많지 않음 (ex, 화씨 -> 섭씨로 표기되는 케이스 없음)

항목 이름	unit_source_value 표기	설명
liter per minute	l/min, l/mi	Unit : liter per minute
beats/min	bpm, bp, b	Unit : beats per minute
breaths/min	breaths/min	Unit : breaths per minute
centimeter	cm	Unit : centimeter
cetimeter watercolumn	cmH2, cmH2O, cmH	Unit : cetimeter watercolumn
degree Celsius	℃	Unit : degree Celsius
kilogram	k, kg	Unit : kilo-gram
L/s	l/sec	Unit : liter per second
liter	l/, l	Unit : liter
liter per minute	l/m	Unit : liter per minute
milliliter	ml	Unit : milliliter
milliliter per minute	ml/min	Unit : milliliter per minute
millimeter	mm	Unit : millimeter
millimeter mercury column	mmHg, mmH	Unit : millimeter mercury column
millivolt	mV	Unit : millivolt
per minute	/min	Unit : per minute
percent	%	Unit : percent
rpm	r, rp, rpm	Unit : rpm
second	s, sec, se	Unit : second
liter per minute	l/min, l/mi	Unit : liter per minute
beats/min	bpm, bp, b	Unit : beats per minute
breaths/min	breaths/min	Unit : breaths per minute
centimeter	cm	Unit : centimeter
cetimeter watercolumn	cmH2, cmH2O, cmH	Unit : cetimeter watercolumn
degree Celsius	℃	Unit : degree Celsius

Challenge Data

OUTCOME_COHORT Table

- 예측 모델에서 예측하여야 할 Cohort 목록
- 예측 모델은 주어진 subject_id(person_id) 환자의 cohort_end_date 시점에서 12시간 이후 critical event 발생 여부를 예측하여야 함
- training set에서는 COHORT 목록에 '**LABEL**' 이라는 정답 필드가 있음

Field	Type	Description
subject_id	integer	예측 대상자의 person_id foreign key
cohort_start_date	datetime	데이터 시작 구간, observation window start datetime (2019-10-10 10:30:00)
cohort_end_date	datetime	데이터 종료 구간, observation window end datetime (2019-11-12 23:00:00)
label	integer	데이터 종료 후 향후 12시간 이내 critical event 발생 여부 (1: critical event, 0: normal)

3. Preprocessing

Challenge Data

SNUH 분당서울대학교병원 SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

데이터 추출 및 가공

- 이미지 데이터와 다르게 예측 모델에 입력될 데이터를 직접 테이블에서 추출하여 가공하여야 함
  - PERSON table -> 환자 성별, 생년월일
  - CONDITION_OCCURRENCE table -> 환자의 진단 및 증상
  - MEASUREMENT table -> 측정된 생체 신호 값
- 의료진의 판단과 중재에 따라 측정되는 생체 신호 항목이 상황에 따라 달라지거나 결측치(missing value), 이상치(outlier)가 발생할 가능성이 있음
  - 1일차 측정 생체 신호 (non-invasive blood pressure, heart rate, respiratory rate) -> 2일차 측정 생체 신호 (arterial blood pressure, heart rate) 과 같은 형식으로 변경
  - respiratory rate (rpm) : 31, 29, 0, 0, 22 ... 와 같은 결측치 발생 가능
- 모델이 주어진 환자 출생시점부터 COHORT_END_DATETIME 까지의 데이터를 분석하여 결과를 추론할 수 있도록 데이터 가공하여야 함

데이터 추출 및 가공 – Person table

- PERSON TABLE에서 환자 정보 조회
  - 환자 고유식별자 (PERSON_ID), 출생일 (BIRTH_DATETIME), 성별 (GENDER_SOURCE_VALUE) 추출한다고 가정
  - 필요에 따라 YEAR_OF_BIRTH, MONTH_OF_BIRTH, DAY_OF_BIRTH 활용 가능

PERSON_ID	GENDER_CONCEPT_ID	YEAR_OF_BIRTH	MONTH_OF_BIRTH	DAY_OF_BIRTH	BIRTH_DATETIME	RACE_CONCEPT_ID	ETHNICITY_CONCEPT_ID	LOCATION_ID	PROVIDER_ID	CARE_SITE_ID	PERSON_SOURCE_VALUE
47715145949628715	8507	2036	4	11	2036-04-11 09:43:00	38003585	0	NaN	NaN	NaN	NaN

GENDER_SOURCE_VALUE	GENDER_SOURCE_CONCEPT_ID	RACE_SOURCE_VALUE	RACE_SOURCE_CONCEPT_ID	ETHNICITY_SOURCE_VALUE	ETHNICITY_SOURCE_CONCEPT_ID
M	0	163	0	NaN	NaN

- PERSON_ID, GENDER, BIRTH_DATETIME 을 추출하여 환자의 demographics 정보 추출

```
person_table[['PERSON_ID','BIRTH_DATETIME','GENDER_SOURCE_VALUE']]
```

	PERSON_ID	BIRTH_DATETIME	GENDER_SOURCE_VALUE
0	47715145949628715	2036-04-11 09:43:00	M

Challenge Data

SNUH 분당서울대학교병원 SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

데이터 추출 및 가공 – Condition_occurrence table

- 대상 환자의 PERSON_ID를 사용하여 condition 정보 조회
- condition_occurrence table 에서 예측에 필요한 필드 확인 및 추출
- CONDITION_SOURCE_VALUE 대신 CONDITION_SOURCE_CONCEPT_ID 를 조회하여 확인 가능

	CONDITION_OCCURRENCE_ID	PERSON_ID	CONDITION_CONCEPT_ID	CONDITION_START_DATE	CONDITION_START_DATETIME	CONDITION_END_DATE	CONDITION_END_DATETIME	CONDITION_TYPE_CONCEPT_ID	STOP_REASON
0	648146764224594	47715145949628715	40489909	2036-04-27 00:00:00	2036-04-27 00:00:00	NaN	NaN	44786629	NaN
1	551970823005861 43	47715145949628715	4079848	2036-04-23 00:00:00	2036-04-23 00:00:00	NaN	NaN	44786629	NaN
2	280117844610741 76	47715145949628715	4283942	2036-04-23 00:00:00	2036-04-23 00:00:00	NaN	NaN	44786629	NaN
3	480789199171975 31	47715145949628715	4048606	2036-04-17 00:00:00	2036-04-17 00:00:00	NaN	NaN	44786629	NaN
4	282717207423275 98	47715145949628715	258866	2036-04-11 00:00:00	2036-04-11 00:00:00	NaN	NaN	44786629	NaN
5	685193193728832 81	47715145949628715	439128	2036-04-11 00:00:00	2036-04-11 00:00:00	NaN	NaN	44786627	NaN
6	251297726239212 43	47715145949628715	4173323	2036-04-11 00:00:00	2036-04-11 00:00:00	NaN	NaN	44786629	NaN
7	455853655640244 39	47715145949628715	4145947	2036-04-11 00:00:00	2036-04-11 00:00:00	NaN	NaN	44786629	NaN
8	357366669055774 47	47715145949628715	4071485	2036-04-11 00:00:00	2036-04-11 00:00:00	NaN	NaN	44786629	NaN
9	670177944951554 21	47715145949628715	433027	2036-04-11 00:00:00	2036-04-11 00:00:00	NaN	NaN	44786629	NaN
10	714080048492486 74	47715145949628715	40482735	2036-04-11 00:00:00	2036-04-11 00:00:00	NaN	NaN	44786629	NaN
11	68427904795401 65	47715145949628715	444104	2036-04-11 00:00:00	2036-04-11 00:00:00	NaN	NaN	42894222	NaN
12	101224651026591 90	47715145949628715	440847	2036-04-12 00:00:00	2036-04-12 00:00:00	NaN	NaN	44786629	NaN

PROVIDER_ID	VISIT_OCCURRENCE_ID	CONDITION_STATUS_CONCEPT_ID	CONDITION_SOURCE_VALUE	CONDITION_SOURCE_CONCEPT_ID	CONDITION_STATUS_SOURCE_VALUE	VISIT_DETAIL_ID
358865415712089 86	7288010611789958	4230359.0	A40.1	45576238.0	Y	NaN
542314206855897 63	7288010611789958	4230359.0	P28.4	45592332.0	Y	NaN
542314206855897 63	7288010611789958	4230359.0	P27.1	45558368.0	Y	NaN
445016163179121 28	7288010611789958	4230359.0	P52.6	45606707.0	Y	NaN
542314206855897 63	7288010611789958	4230359.0	P22.0	45544055.0	Y	NaN
643215838251769 38	7288010611789958	4230359.0	P07.2	45577689.0	Y	NaN
643215838251769 38	7288010611789958	4230359.0	P07.0	45597065.0	Y	NaN
542314206855897 63	7288010611789958	4230359.0	P05.1	45572924.0	Y	NaN
542314206855897 63	7288010611789958	4230359.0	P03.4	45558357.0	Y	NaN
542314206855897 63	7288010611789958	4230359.0	P00.0	45577680.0	Y	NaN
643215838251769 38	7288010611789958	4230359.0	Z38.0	45566424.0	Y	NaN
643215838251769 38	7288010611789958	NaN	NaN	NaN	NaN	NaN
542314206855897 63	7288010611789958	4230359.0	P59.0	45597088.0	Y	NaN

데이터 추출 및 가공 – Condition_occurrence table

- 해당 테이블의 PERSON_ID, CONDITION_START_DATETIME, CONDITION_SOURCE_VALUE로 시간대별 진단 및 증상 정보 확인 가능
  - CONDITION_CONCEPT_ID 로 구체적인 정보 확인 가능

	PERSON_ID	CONDITION_START_DATETIME	CONDITION_SOURCE_VALUE
0	47715145949628715	2036-04-27 00:00:00	A40.1
1	47715145949628715	2036-04-23 00:00:00	P28.4
2	47715145949628715	2036-04-23 00:00:00	P27.1
3	47715145949628715	2036-04-17 00:00:00	P52.6
4	47715145949628715	2036-04-11 00:00:00	P22.0
5	47715145949628715	2036-04-11 00:00:00	P07.2
6	47715145949628715	2036-04-11 00:00:00	P07.0
7	47715145949628715	2036-04-11 00:00:00	P05.1
8	47715145949628715	2036-04-11 00:00:00	P03.4
9	47715145949628715	2036-04-11 00:00:00	P00.0
10	47715145949628715	2036-04-11 00:00:00	Z38.0
11	47715145949628715	2036-04-11 00:00:00	NaN
12	47715145949628715	2036-04-12 00:00:00	P59.0

	PERSON_ID	CONDITION_START_DATETIME	CONDITION_CONCEPT_ID	concept_name
0	47715145949628715	2036-04-27 00:00:00	40489909	Sepsis due to Streptococcus agalactiae
1	47715145949628715	2036-04-23 00:00:00	4079848	Apnea of prematurity
2	47715145949628715	2036-04-23 00:00:00	4283942	Bronchopulmonary dysplasia of newborn
3	47715145949628715	2036-04-17 00:00:00	4048606	Cerebellar (nontraumatic) and posterior fossa ...
4	47715145949628715	2036-04-11 00:00:00	258866	Respiratory distress syndrome in the newborn
5	47715145949628715	2036-04-11 00:00:00	439128	Extreme prematurity of infant
6	47715145949628715	2036-04-11 00:00:00	4173323	Extremely low birth weight infant
7	47715145949628715	2036-04-11 00:00:00	4145947	Small for gestational age fetus
8	47715145949628715	2036-04-11 00:00:00	4071485	Fetal or neonatal effect of cesarean section
9	47715145949628715	2036-04-11 00:00:00	433027	Fetal or neonatal effect of maternal hypertens...
10	47715145949628715	2036-04-11 00:00:00	40482735	Liveborn born in hospital
11	47715145949628715	2036-04-11 00:00:00	444104	Newborn
12	47715145949628715	2036-04-12 00:00:00	440847	Neonatal jaundice associated with preterm deli...

Challenge Data

SNUH 분당서울대학교병원 SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

데이터 추출 및 가공 – Measurement table

- patient monitor에서 생성된 생체신호 정보의 항목, 시간, 측정치 조회
- 측정 일시와 측정 항목 이름, 단위, 수치형 측청값을 추출

	MEASUREMENT_ID	PERSON_ID	MEASUREMENT_CONCEPT_ID	MEASUREMENT_DATE	MEASUREMENT_DATETIME	MEASUREMENT_TYPE_CONCEPT_ID	OPERATOR_CONCEPT_ID	VALUE_AS_NUMBER	VALUE_AS_CONCEPT_ID
0	477151459496287 16	477151459496287 15	4239408.0	2036-04-22 00:00:00	2036-04-22 23:34:00	44818704	NaN	163.0	NaN
1	477151459496287 17	477151459496287 15	4302666.0	2036-04-22 00:00:00	2036-04-22 23:34:00	44818704	NaN	36.8	NaN
2	477151459496287 18	477151459496287 15	4313591.0	2036-04-22 00:00:00	2036-04-22 23:34:00	44818704	NaN	49.0	NaN
3	477151459496287 19	477151459496287 15	4011919.0	2036-04-22 00:00:00	2036-04-22 23:34:00	44818704	NaN	96.0	NaN
4	477151459496287 20	477151459496287 15	4239408.0	2036-04-22 00:00:00	2036-04-22 23:34:00	44818704	NaN	162.0	NaN
5	477151459496287 21	477151459496287 15	4011919.0	2036-04-22 00:00:00	2036-04-22 23:34:30	44818704	NaN	96.0	NaN
6	477151459496287 22	477151459496287 15	4302666.0	2036-04-22 00:00:00	2036-04-22 23:34:30	44818704	NaN	36.8	NaN
7	477151459496287 23	477151459496287 15	4239408.0	2036-04-22 00:00:00	2036-04-22 23:34:30	44818704	NaN	164.0	NaN
8	477151459496287 24	477151459496287 15	4313591.0	2036-04-22 00:00:00	2036-04-22 23:34:30	44818704	NaN	53.0	NaN
9	477151459496287 25	477151459496287 15	4239408.0	2036-04-22 00:00:00	2036-04-22 23:34:30	44818704	NaN	164.0	NaN

UNIT_CONCEPT_ID	RANGE_LOW	RANGE_HIGH	PROVIDER_ID	VISIT_OCCURRENCE_ID	MEASUREMENT_SOURCE_VALUE	MEASUREMENT_SOURCE_CONCEPT_ID	UNIT_SOURCE_VALUE	VALUE_SOURCE_VALUE	VISIT_DETAIL_ID
4118124.0	NaN	NaN	5.018011e+16	7.288011e+15	HR	40767007	bpm	163.0	NaN
NaN	NaN	NaN	5.018011e+16	7.288011e+15	Temp	40767007	°C	36.8	NaN
4121488.0	NaN	NaN	5.018011e+16	7.288011e+15	RR	40767007	rpm	49.0	NaN
8554.0	NaN	NaN	5.018011e+16	7.288011e+15	SpO2	40767007	%	96.0	NaN
4118124.0	NaN	NaN	5.018011e+16	7.288011e+15	Pulse	40767007	bpm	162.0	NaN
8554.0	NaN	NaN	5.018011e+16	7.288011e+15	SpO2	40767007	%	96.0	NaN
NaN	NaN	NaN	5.018011e+16	7.288011e+15	Temp	40767007	°C	36.8	NaN
4118124.0	NaN	NaN	5.018011e+16	7.288011e+15	HR	40767007	bpm	164.0	NaN
4121488.0	NaN	NaN	5.018011e+16	7.288011e+15	RR	40767007	rpm	53.0	NaN
4118124.0	NaN	NaN	5.018011e+16	7.288011e+15	Pulse	40767007	bpm	164.0	NaN

데이터 추출 및 가공 – Measurement table

- patient monitor에서 생성된 생체신호 정보의 항목, 시간, 측정치 조회
- 측정 일시와 측정 항목 이름, 단위, 수치형 측청값을 추출
- 측정한 항목에서 이상치값에 대한 처리 수행

	PERSON_ID	MEASUREMENT_DATETIME	VALUE_AS_NUMBER	MEASUREMENT_SOURCE_VALUE	UNIT_SOURCE_VALUE
0	47715145949628715	2036-04-11 10:04:00	93.0	SpO2	%
1	47715145949628715	2036-04-11 10:04:00	0.0	Weight	kg
2	47715145949628715	2036-04-11 10:04:00	0.0	Height	cm
3	47715145949628715	2036-04-11 10:04:00	159.0	Pulse	bpm
4	47715145949628715	2036-04-11 10:04:30	93.0	SpO2	%
5	47715145949628715	2036-04-11 10:04:30	158.0	Pulse	bpm
6	47715145949628715	2036-04-11 10:05:00	157.0	Pulse	bpm
7	47715145949628715	2036-04-11 10:05:00	93.0	SpO2	%
8	47715145949628715	2036-04-11 10:05:30	93.0	SpO2	%
9	47715145949628715	2036-04-11 10:05:30	157.0	Pulse	bpm

데이터 추출 및 가공 – Measurement table

- patient monitor에서 생성된 생체신호 정보의 항목, 시간, 측정치 조회
- 측정 일시와 측정 항목 이름, 단위, 수치형 측청값을 추출
- 측정한 항목에서 이상치값에 대한 처리 수행

	PERSON_ID	MEASUREMENT_DATETIME	VALUE_AS_NUMBER	MEASUREMENT_SOURCE_VALUE	UNIT_SOURCE_VALUE
0	47715145949628715	2036-04-11 10:04:00	93.0	SpO2	%
1	47715145949628715	2036-04-11 10:04:00	0.0	Weight	kg
2	47715145949628715	2036-04-11 10:04:00	0.0	Height	cm
3	47715145949628715	2036-04-11 10:04:00	159.0	Pulse	bpm
4	47715145949628715	2036-04-11 10:04:30	93.0	SpO2	%
5	47715145949628715	2036-04-11 10:04:30	158.0	Pulse	bpm
6	47715145949628715	2036-04-11 10:05:00	157.0	Pulse	bpm
7	47715145949628715	2036-04-11 10:05:00	93.0	SpO2	%
8	47715145949628715	2036-04-11 10:05:30	93.0	SpO2	%
9	47715145949628715	2036-04-11 10:05:30	157.0	Pulse	bpm

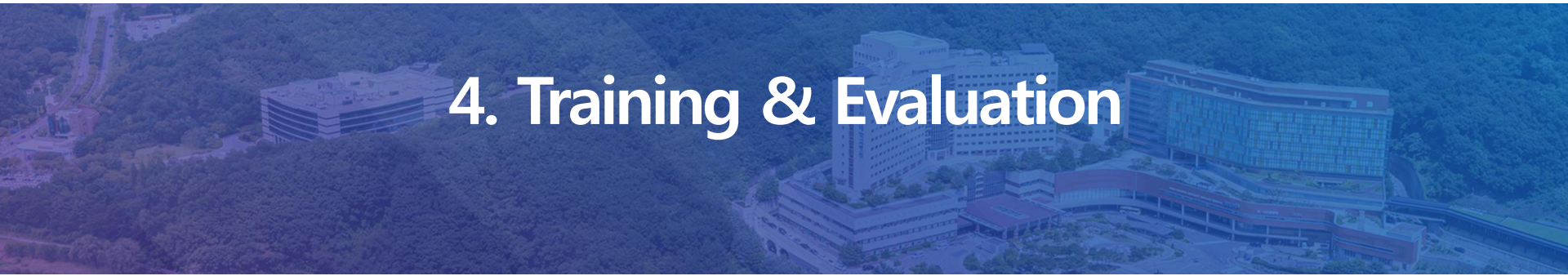
Challenge Data

SNUH 분당서울대학교병원 SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

데이터 가공

- 데이터 추출 완료 후 OUTCOME_COHORT 의 정보를 기준으로 모델 입력용 데이터셋 구축
- 아래는 데이터셋 구축 예시

```
x :  [8.339997335464536, -1.5, 47.5, 22.0, 21.0, 8.73053390247253, 21.0, 191.0, 7.176995623860602, 148.0, 150.4396671289875, 0.0, 99.15715121719043, 1.277741179662437, 9.0, 0, 0, 0, 0, 10.713333291001529]
y :  0
x :  [9.5, -2.0, 47.333333333333336, 22.0, 21.0, 9.5, 21.0, 191.0, 7.948643267699245, 148.0, 149.71775312066575, 0.0, 99.18489046823065, 1.123278135196876, 9.0, 0, 0, 0, 0, 11.116804097101529]
y :  0
x :  [0.0, -3.0, 50.0, 22.0, 21.0, 0.0, 21.0, 180.0, 7.3708355526564695, 148.0, 150.83891660727014, 0.0, 99.36194563662374, 0.9181268727013718, 9.0, 0, 0, 0, 0, 10.016652800877813]
y :  0
x :  [7.1336448530109, -1.4999999999999996, 42.75, 34.0, 38.0, 8.806563209081938, 21.0, 179.0, 5.718261367992508, 148.0, 150.18700927908637, 0.0, 99.06886657101865, 0.7674068335096584, 4.0, 0, 0, 0, 0, 10.016652800877813]
y :  0
x :  [7.1336448530109, -1.4999999999999996, 42.75, 34.0, 38.0, 8.806563209081938, 21.0, 197.0, 5.2254385317048l9, 151.0, 152.8893647394718, 0.0, 98.79913916786226, 0.8246763478736602, 5.0, 0, 0, 0, 0, 10.656032113698009]
y :  0
x :  [12.283683848458853, -1.5, 43.0, 34.0, 38.0, 14.72714802291635, 18.0, 197.0, 5.490585060094l3, 150.0, 152.38115631691647, 0.0, 98.52367288378767, 0.8250229701699018, 5.0, 0, 0, 0, 0, 10.935416468216166]
y :  0
x :  [5.557777333511022, -1.5, 36.0, 17.0, 18.0, 6.79869268479038, 18.0, 197.0, 4.8146204297624955, 151.0, 153.33404710920772, 0.0, 98.25538020086083, 0.8767417502196115, 6.0, 0, 0, 0, 0, 5.343739847293799]
y :  0
x :  [5.557777333511022, -1.5, 36.0, 17.0, 18.0, 6.79869268479038, 18.0, 197.0, 5.220773860109306, 151.0, 153.1220556745182, 0.0, 98.06025824964132, 0.8745256627224876, 6.0, 0, 0, 0, 0, 5.343739847293799]
y :  0
x :  [4.949747468305833, -1.843210323827269, 37.8, 17.0, 18.0, 6.49519052838329, 18.0, 197.0, 4.780443834291457, 151.0, 154.14315496872828, 0.0, 97.76044568245125, 0.9273801609862289, 8.0, 0, 0, 0, 0, 5.421047417431508]
y :  0
x :  [2.5, -2.0, 39.333333333333336, 12.0, 15.0, 4.5, 15.0, 197.0, 4.341543778268949, 153.0, 155.5350451075642, 0.0, 97.99583333333334, 0.9708545357365628, 8.0, 0, 0, 0, 0, 4.707440918375928]
y :  0
```



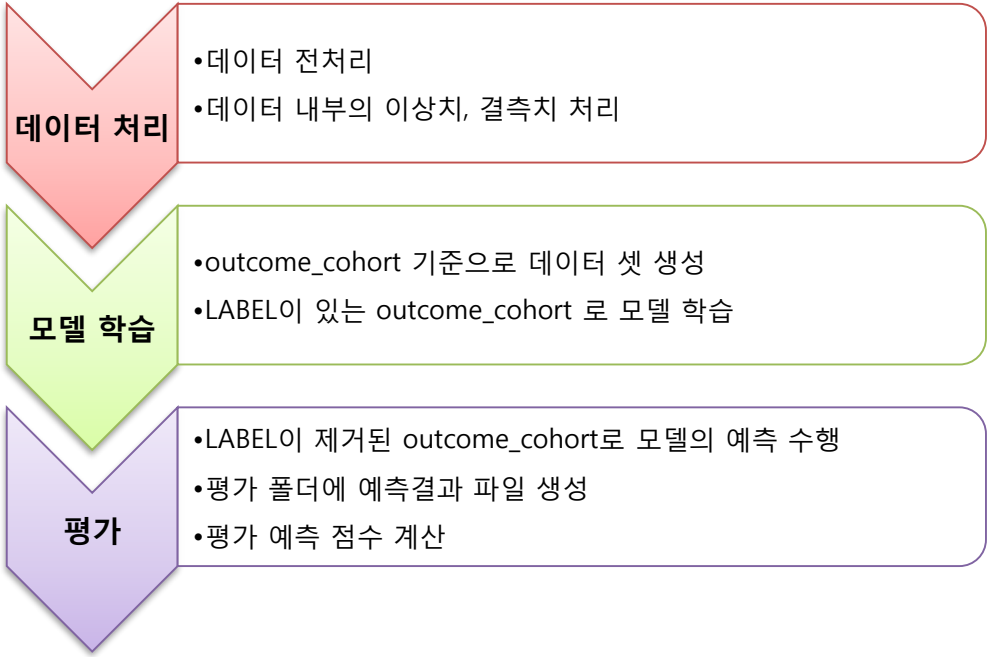
4. Training & Evaluation

Training & Evaluation

SNUH 분당서울대학교병원 SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

Training / Evaluation 절차

- 모델의 학습 및 평가 절차는 OUTCOME_COHORT를 중심으로 수행
- LABEL 은 정답셋
    - 1 : event 발생
    - 0 : event 발생하지 않음
    - 모델 학습 시에만 조회 가능
- Training에서는 시간 제약 없이 전체 항목에 대해 조회 및 학습 가능
- evaluation 수행 시 모든 테이블은 OUTCOME_COHORT 에 기재된 START_DATE과 END_DATE 까지의 시간대만 조회 가능



Training & Evaluation

Training – outcome cohort 조회 및 label 확인

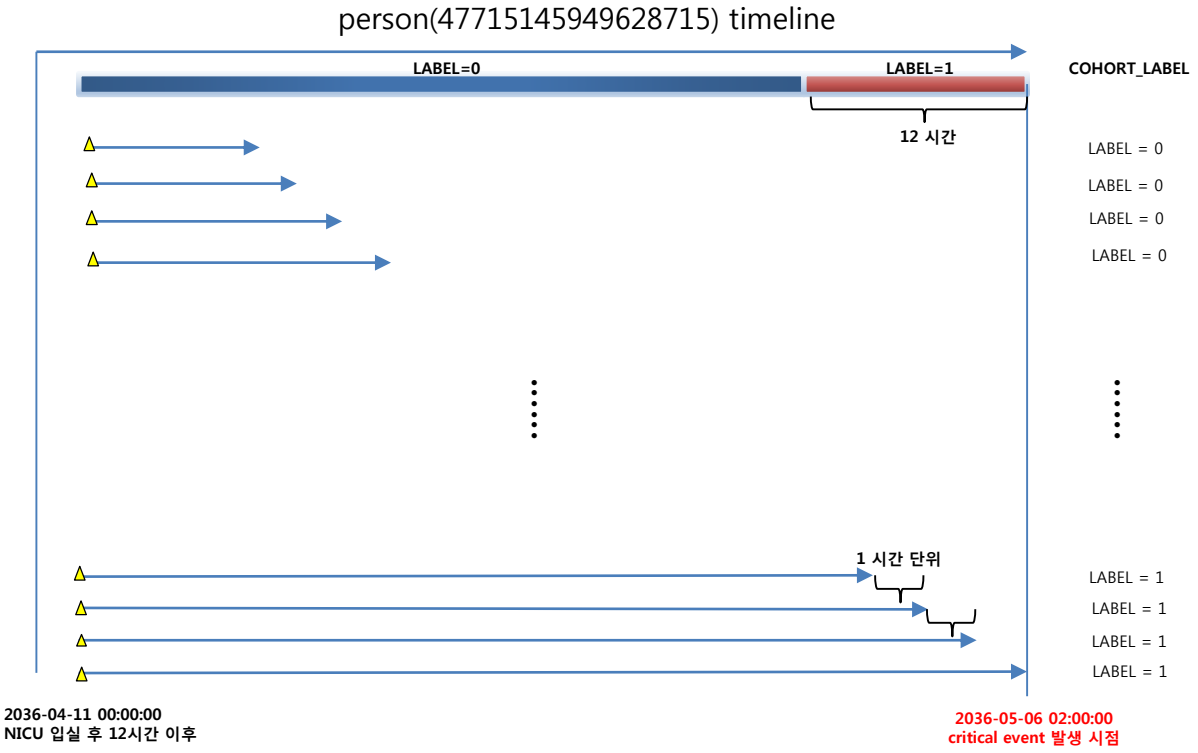
- 예측 모델은 COHORT_END_DATE 시점에서 12시간 이내에 환자가 critical event가 발생할지 여부에 대해 추론하여야 함
- LABEL 은 정답셋
    - 1 : event 발생
    - 0 : event 발생하지 않음
- Training에서는 시간 제약 없이 전체 항목에 대해 조회 및 학습 가능

COHORT_DEFINITION_ID	SUBJECT_ID	COHORT_START_DATE	COHORT_END_DATE	LABEL
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 12:00:00	0
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 13:00:00	0
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 14:00:00	0
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 15:00:00	1
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 16:00:00	1
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 17:00:00	1
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 18:00:00	1
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 19:00:00	1
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 20:00:00	1
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 21:00:00	1
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 22:00:00	1
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 23:00:00	1
1	47715145949628715	2036-04-11 00:00:00	2036-05-06 00:00:00	1
1	47715145949628715	2036-04-11 00:00:00	2036-05-06 01:00:00	1
1	47715145949628715	2036-04-11 00:00:00	2036-05-06 02:00:00	1

Training & Evaluation

Training – outcome cohort 조회 및 label 확인

- 예측 모델은 COHORT_END_DATE 시점에서 12시간 이내에 환자가 critical event가 발생할지 여부에 대해 추론하여야 함



Training & Evaluation

Evaluation – evaluation 진행 과정

- LABEL field는 삭제
- 예측 모델은 test set에 있는 person table, measurement table, condition_occurrence table, outcome_cohort table을 입력으로 받음
- 모델은 주어진 outcome_cohort 순서대로 LABEL과 LABEL_PROBABILITY 를 기록한 정답데이터를 출력

LABEL 열 삭제

COHORT_DEFINITION_ID	SUBJECT_ID	COHORT_START_DATE	COHORT_END_DATE
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 12:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 13:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 14:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 15:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 16:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 17:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 18:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 19:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 20:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 21:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 22:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-05 23:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-06 00:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-06 01:00:00
1	47715145949628715	2036-04-11 00:00:00	2036-05-06 02:00:00

Training & Evaluation

Evaluation 예시 – 결과 출력

- 모델은 주어진 OUTCOME_COHORT의 입력에 대응하여 다음과 같은 결과를 출력해야 함
  - LABEL
  - LABEL_PROBABILITY

모델 출력 파일(csv)

LABEL	LABEL_PROBABILITY
1	1
1	1
1	0.86
1	0.77
0	0
0	0.4
1	0.8
1	0.65
1	0.66
1	0.76
0	0.22
0	0.31
0	0.44
0	0.21
0	0

Training & Evaluation

SNUH 분당서울대학교병원 SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

Evaluation – 점수 계산

- 점수는 실제 정답 LABEL 데이터와 비교하여 계산
- 계산에 활용되는 metrix
  - Area under the curve receiver operating characteristics (AUROC)
  - F1-score
- 최종적으로는 두 정량값의 harmonic mean으로 계산
  - 2 * (AUROC * F1) / (AUROC + F1)
  - 옆의 식의 결과는 0.6416243654822336

정답 LABEL

LABEL
1
1
1
1
1
1
0
0
0
0
0
0
0
0
0

모델 출력 파일(csv)

LABEL	LABEL_PROBABILITY
1	1
1	1
1	0.86
1	0.77
0	0
0	0.4
1	0.8
1	0.65
1	0.66
1	0.76
0	0.22
0	0.31
0	0.44
0	0.21
0	0

Training & Evaluation

SNUH 분당서울대학교병원 SEOUL NATIONAL UNIVERSITY BUNDANG HOSPITAL

데이터 분류 및 비율

- 데이터는 Training, validation, test가 6:2:2 으로 분류
- 각 데이터 셋 간의 동일한 환자는 없으므로 환자 식별자로 학습하더라도 결과에 영향을 주지 않음
- NICU 생체 신호 training set 의 positive/negative case 비율
    - positive : 506 건 (0.59 %)
    - negative : 84759 건 (99.41 %)
- Negative case로만 대부분 선택해도 높은 수준의 ROC 를 얻을 수 있을 가능성이 있기에 F1-measurement의 harmonic mean이 추가적 정량평가에 포함됨
- 따라서 Positive predictive value(PPV) 가 높을 수록 높은 점수를 얻을 확률이 큼