

An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU

2021312101 DA 문구영

문제의식

① 패혈증 조기 진단의 필요성

- 패혈증 : 중환자실 환자들에게 있어 병적 병리, 사망, 높은 경제적 비용의 주된 원인
- 패혈증 임상 징후에 앞서 조기 적절한 항생제 치료가 사망률, 비용 절감에 있어 중요
But 임상적으로 검증된 패혈증 발병 **실시간 예측 시스템**의 부재

② 기존 연구들이 사용한 Dataset의 한계

- EMR, Laboratory data : **static (low resolution)** & 낮고, 일관적이지 않은 빈도로 수집 ex) 성별, 연령, 인종
→ **dynamic (high resolution, real time)** data 활용의 필요성 ex) 심박수, 혈압
- **Dynamic data** during hypotensive episode improves mortality predictions among patients with sepsis and hypotension. Crit Care Med 2013; 41:954–962

연구목적

- ① T 시간($T=12, 8, 6, 4$) 이내의 패혈증 발병을 예측하는 **AISE(Artificial Intelligence Sepsis Expert)** 알고리즘 개발, 검증
- ② **Low resolution EMR & High resolution 생리학 Data**를 결합하여 고성능의 예측 모델 구축, 검증
- ③ AISE 알고리즘의 **generalizability & interpretability** 확인
- ④ 패혈증 발병 예측에 대한 기여도가 높은 특성 변수(**contributing factor**) 파악

Methods

Cohort

- **Source**

- ① **Development Cohort** : Emory University Hospital ICU (약 31000건)
- ② **Validation Cohort** : MIMIC 3 DB (약 52000건)

Feature Extraction & Measurements

- EMR Data, 고빈도의 생리학 Data에서 65개의 특성 변수 추출
- 연속형 변수 : 특정 기간 동안 측정된 값들의 **median**, 양측 Wilcoxon rank-sum 검정 (septic vs non-septic)
- 범주형 변수 : **비율**, 양측 카이제곱 검정 (septic vs non-septic)

- * **Wilcoxon rank-sum**

- septic / non-septic 그룹의 data가 중앙값이 같은 연속 분포에서 추출된 표본인지 검정 (비모수적)

Methods

Research Hypothesis

1. What is your EXACT question?

→ 환자의 패혈증 발병 : vague / ICU 퇴원 혹은 패혈증 발병 까지 관찰된 환자가 'Sepsis-3' 기준을 만족 : BETTER

2. What data will you need?

→ ICU 환자 : vague / 18세 이상의 ICU 입원 환자 중, 체류 기간이 8시간 이상 20일 미만이며, ICU 입실 후 4시간 이내로 패혈증이 발병 되지 않은 환자 : BETTER

→ 혈압 : vague / 특정 시간 내 측정된 혈압의 중앙값 : BETTER – **Feature Measurements**

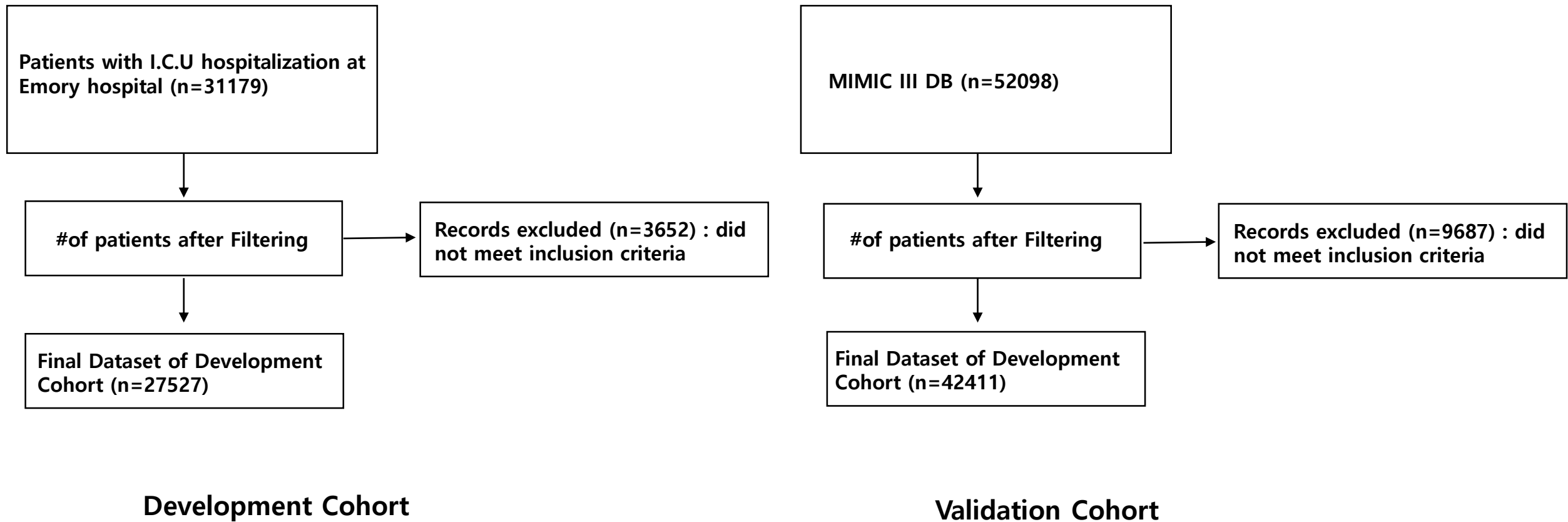
3. Is your question QUANTIFIABLE with a statistic test?

→ 변수 A가 패혈증에 해롭다 : bad / 변수 A가 B보다 패혈증 발병에 대한 **contributing factor**가 높다 - **Interpretability**

4. Is the data collected during ROUTINE care?

→ 시계열 특성 변수는 1시간 마다 update – **Feature Measurements**

Methods – Selection Flow of Cohort



Methods –Features & Measurements

Static features + Dynamic features (clinical & laboratory)

- ① **Dynamic** (11) : RR**STD** / MAP**STD** / **AVG** of multiscale entropy of RR
- ② **Clinical** (10) : MAP / HR / O2Sat / SBP / DBP / RESP / TEMP / GCS / ...
- ③ **Laboratory** (25) : WBC / Hemoglobin / Hematocrit / Creatinine / PTT / INR / ...
- ④ **Demographic** (19) : 상처의 타입 / 수술 유형 / CCI / SOFA 점수 변동

Dynamic 변수들의 측정

- 1시간 단위의 **중복되지 않는 시계열 구간**에서 측정 (1시간 단위로 update) → 측정값들 사이의 선형 관계 (자기상관) 방지
- 1시간 이내에 측정이 여러 번 된 경우 중앙값 사용 – 이상치에 강건
- Mean Imputation으로 결측치 대체

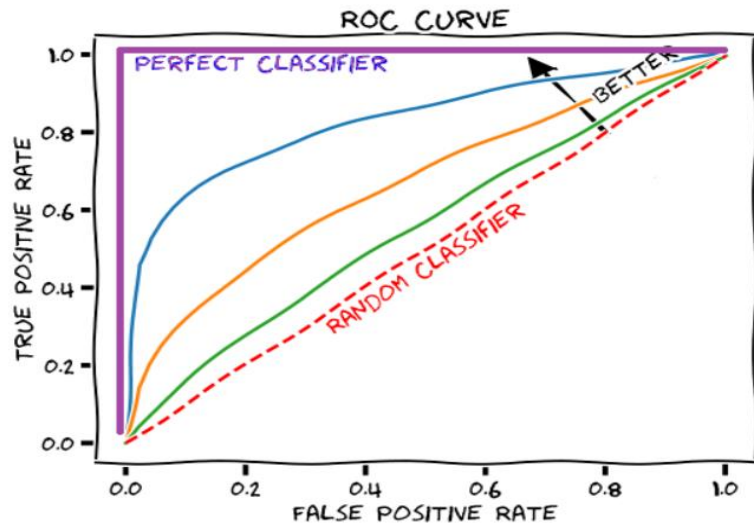
* 저빈도로 측정된 clinical 변수를 제외한 reduced model (53 features)과 full model (65 features)의 예측 성능은 유사

Methods –Features & Measurements

Target Variables

- t_{sepsis} : 패혈증 발병 시간 - ($t_{suspicion} - 24\text{ hr}$, $t_{suspicion} + 12\text{ hr}$) 구간 동안 SOFA 점수가 2점 이상 변동
 - * $t_{suspicion}$: 감염 의심 증상의 *timestamp* – min(항생제 투여, 혈액 배양)
- t_{SOFA} : Sequential Organ Failure Assessment score
- t_{onset} : $\min(t_{SOFA}, t_{sepsis})$

Metrics

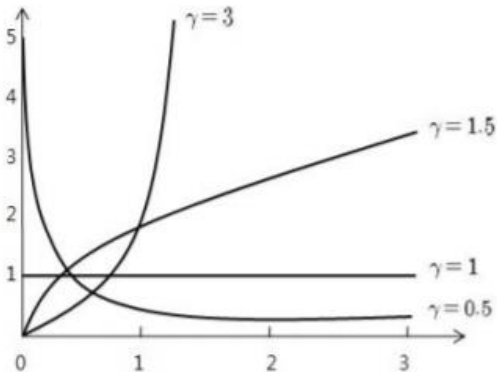


- TP, FP 사용
- 다양한 분류 임계값의 TPR, FPR을 표현

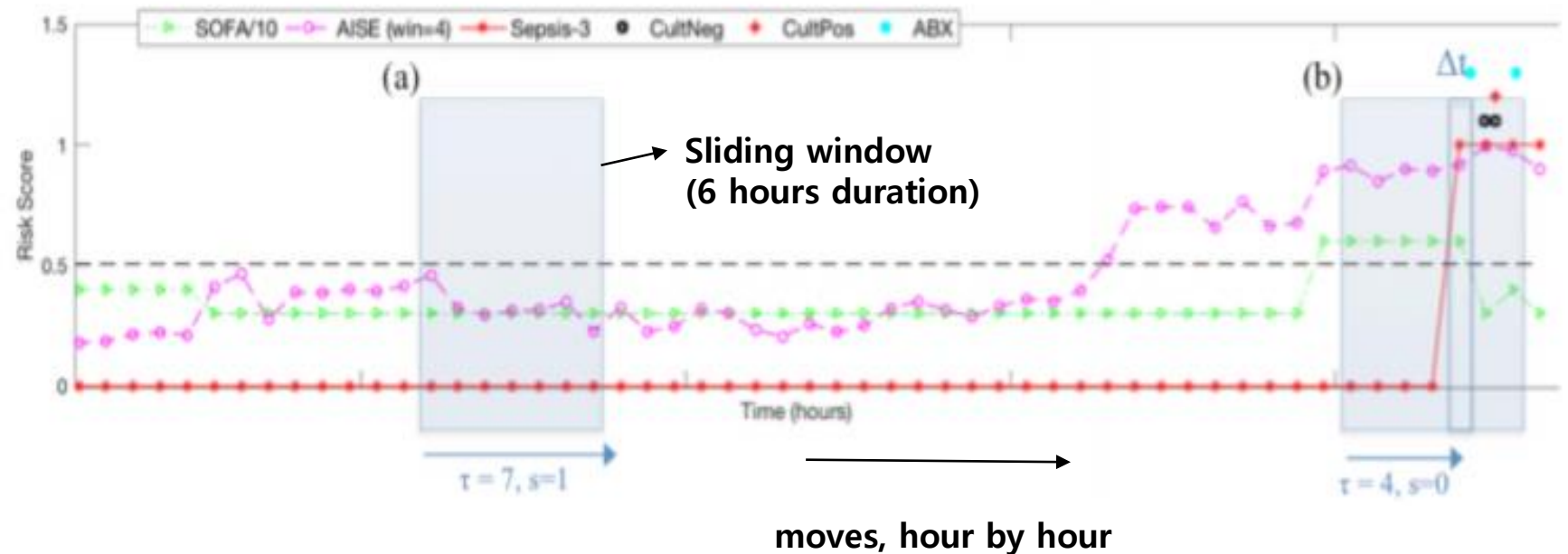
Model

Modified Weibull-Cox Proportional Hazards Model

- T 시간(12, 8, 6, 4) 이내에 패혈증의 발병을 예측하는 생존분석 모델
- ① 기존의 Cox 비례 위험 모형 + ② Weibull 기반 hazard rate
 - ① 모형에 근거하여 회귀계수 추정, 생존함수의 분포에 대한 가정 X : semi-parametric → **robust, generalizability ↑**
 - ② 비례적 위험 함수 가정 (↔ 특정 변수의 위험비는 시간에 상관없이 일정), 로그-선형 모형 → 다양한 패턴 포함



<그림3> Weibull분포의 위험함수 ($\lambda = 1$ 의 경우)



Model

Terms

- ① 생존함수 $S(T)$ (Survival Function) : 환자가 특정 시간 T보다 오래 생존할 확률 (패혈증 미발생)
- ② 위험함수 $h(T)$ (Hazard Function) : T 시점 전까지 생존, 정확히 T 시점에 패혈증이 발생할 확률
- baseline risk h_0 : 기저 위험 + patient specific risk e^{BX} : 환자가 보유한 특성 변수들의 함수
- ③ 누적 위험함수 $\Lambda_0(T)$ (Cumulative Hazard Function) : T 시점 전까지 패혈증이 발생할 확률 합산 (위험함수 0~T 적분)
- ④ 위험률 (Hazard Rate) : T 시점까지 생존한 환자가, 그 직후 패혈증이 발병할 확률 $h_i(\tau|x_i, \lambda, k, \beta) = h_0(\tau|\lambda, k)e^{\beta x_i}$,
product of baseline risk & patient specific risk
- ⑤ 생존 그래프 (Survival Curve) : T 시점까지의 생존 확률 $S(\tau|x_i, \lambda, k, \beta) = e^{-\Lambda_0(\tau)e^{\beta x_i}}$

Model

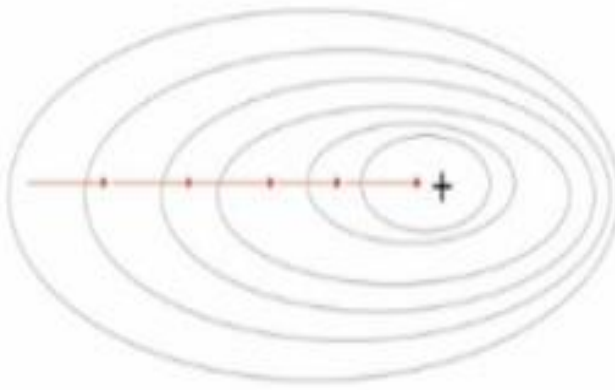
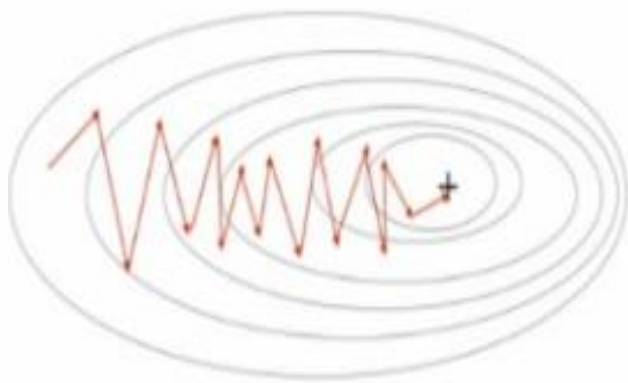
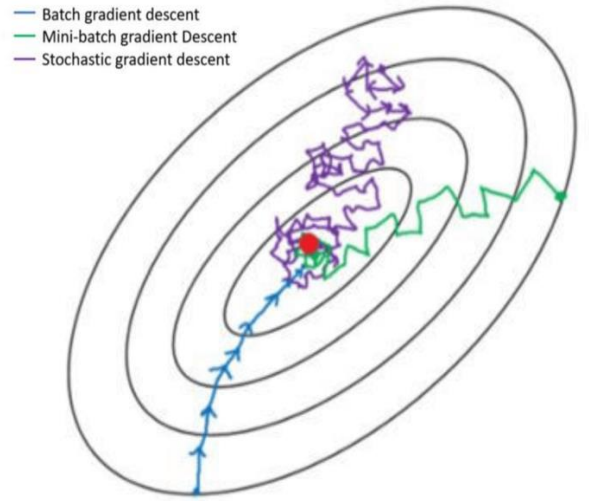
Sepsis Risk Score

- T 시간 이내에 패혈증이 발생할 확률 $\text{Prob}(t_{\text{sepsis}} \leq T) = 1 - S(T|x_i, \lambda, k,)$
- 일 단위 예측보다 유의미한 결과 제공 - A targeted **real-time early warning score** (TREWScore) for septic shock. Sci Transl Med 2015. 7(299)
- Dataset 내에서 무수히 많은 T 시간 단위 **예측 window**가 생성됨 → class 불균형 문제

Model Parameter 학습 ① MSGD(Mini Batch Stochastic Gradient Descent)

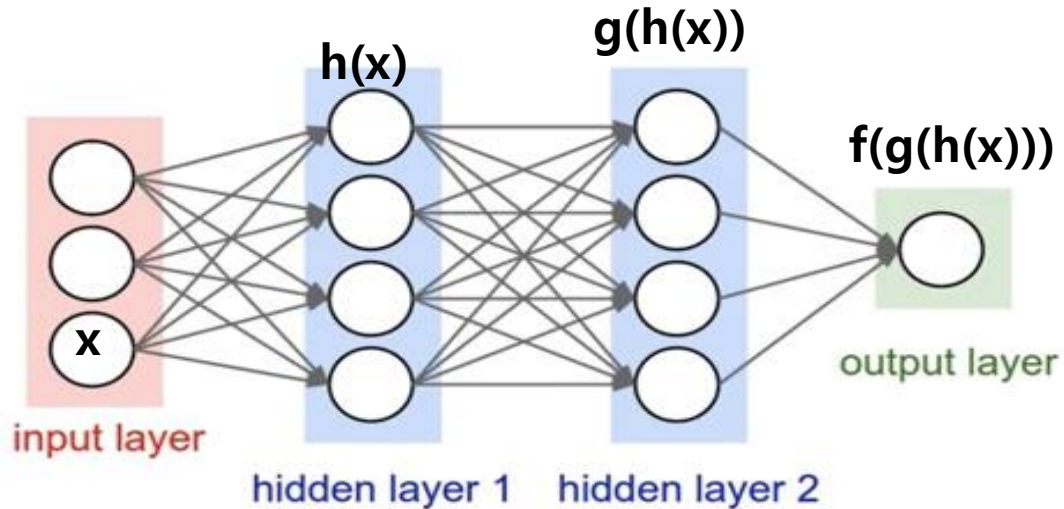
1. 전체 Dataset에서 m개의 data를 추출해 **mini-batch** 구성
2. mini-batch내 data들의 평균 기울기를 통해 parameter update
→ **SGD의 노이즈를 줄이면서, 전체 배치(GD)보다는 빠르게 최적점에 도달**
3. 대용량의 학습 데이터에 확장 가능, 각 mini batch 내에서의 under-sampling을 통해 class 불균형 해소

Appendix – Gradient Descent Methods

	GD	SGD	MSGD
1회 학습에 사용되는 data	모든 data (전체 batch)	랜덤으로 추출된 1개의 data (하나씩 최소화)	전체 표본n을 k개의 mini batch로 분할
학습 방식 (1 epoch)	모든 데이터를 계산하여 최적의 한 step	랜덤하게 추출된 하나의 data만 계산 (속도 ↑)	모수를 k번 최신화, 반복 → 전체 표본 n에 대해 최신화
특징	노이즈 없음 but 국소해 문제	국소해 문제 해결 but 노이즈 심함 (수렴 방향이 임의로 움직임)	국소해 문제 해결 + 노이즈 완화 + mini batch를 통해 시간 단축
해를 찾는 과정	<p>Gradient Descent</p> 	<p>Stochastic Gradient Descent</p> 	 <p>— Batch gradient descent — Mini-batch gradient Descent — Stochastic gradient descent</p>

Appendix – Back Propagation

Chain Rule



→ 학습 목표 : 예측에 대한 손실 함수를 최소화 (편미분값=0)

- Input layer : x
- Hidden layer 1 : $h(x)$
- Hidden layer 2 : $g(h(x))$
- **Output : $f(g(h(x)))$**

$\{f(g(x))\}' = f'(g(x)) \times g'(x)$ → 모델이 복잡할 경우 개별 parameter의 편미분 값을 계산하기 어려움

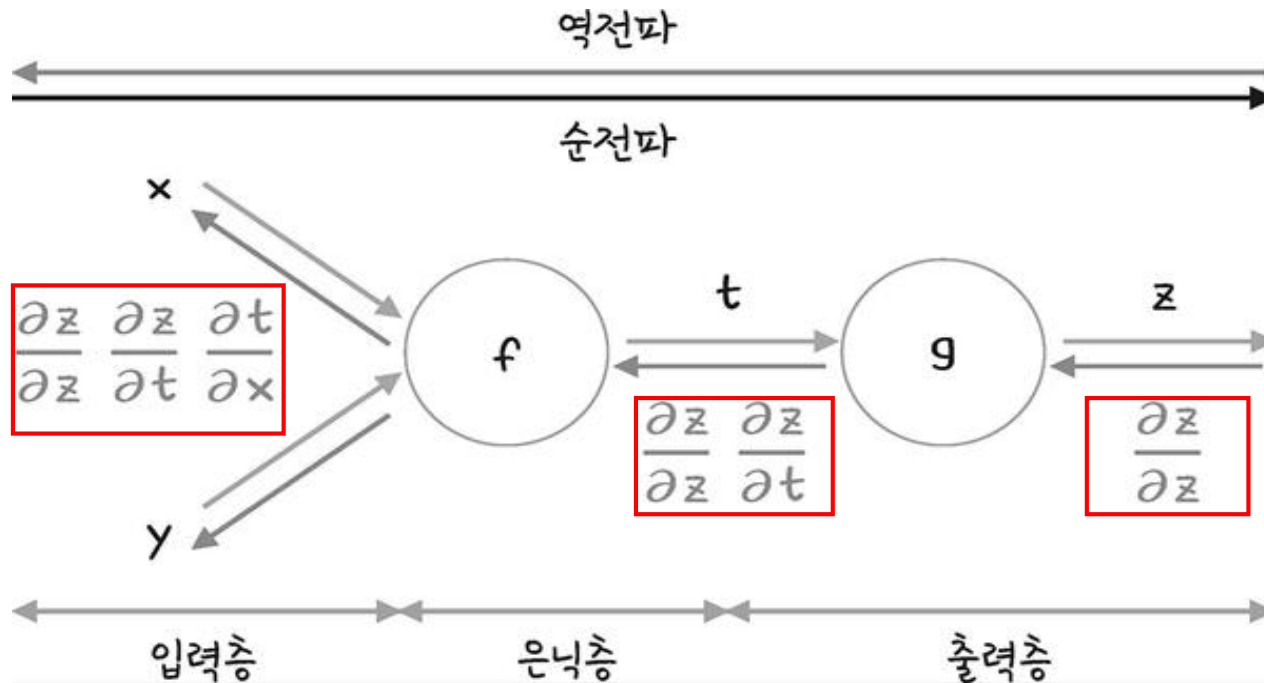
$$\frac{df}{dx} = \frac{df}{dg} \cdot \frac{dg}{dh} \cdot \frac{dh}{dx}$$

- Chain Rule 이용하여 연속된 함수의 미분을 각각의 미분의 곱으로 단순화
- 국소적 계산을 전파하여 최종 결과를 얻음

Appendix – Back Propagation

Back Propagation

- 신경망 모델이 아니어도, 미분 가능한 모델을 사용하는 경우 특정 parameter가 변했을 때 예측 결과가 변화하는 정도를 **back propagation**을 통해 계산 가능
- 매개 변수가 많고, 모델이 복잡해도 효율적인 계산 가능 (경로 사이에 존재하는 모든 국소 미분 값을 곱함)



Model

Model Parameter 학습 ② Back Propagation

- Sepsis Risk Score 변화(미분)에 대한 각 입력 변수들의 **time-varying contribution**을 정량화
→ AISE 알고리즘의 투명성과 **interpretability** ↑
- **Chain Rule**을 이용하여 Model parameter(가중치)와 입력의 변화에 따른 Sepsis Risk Score 변화 정도 계산

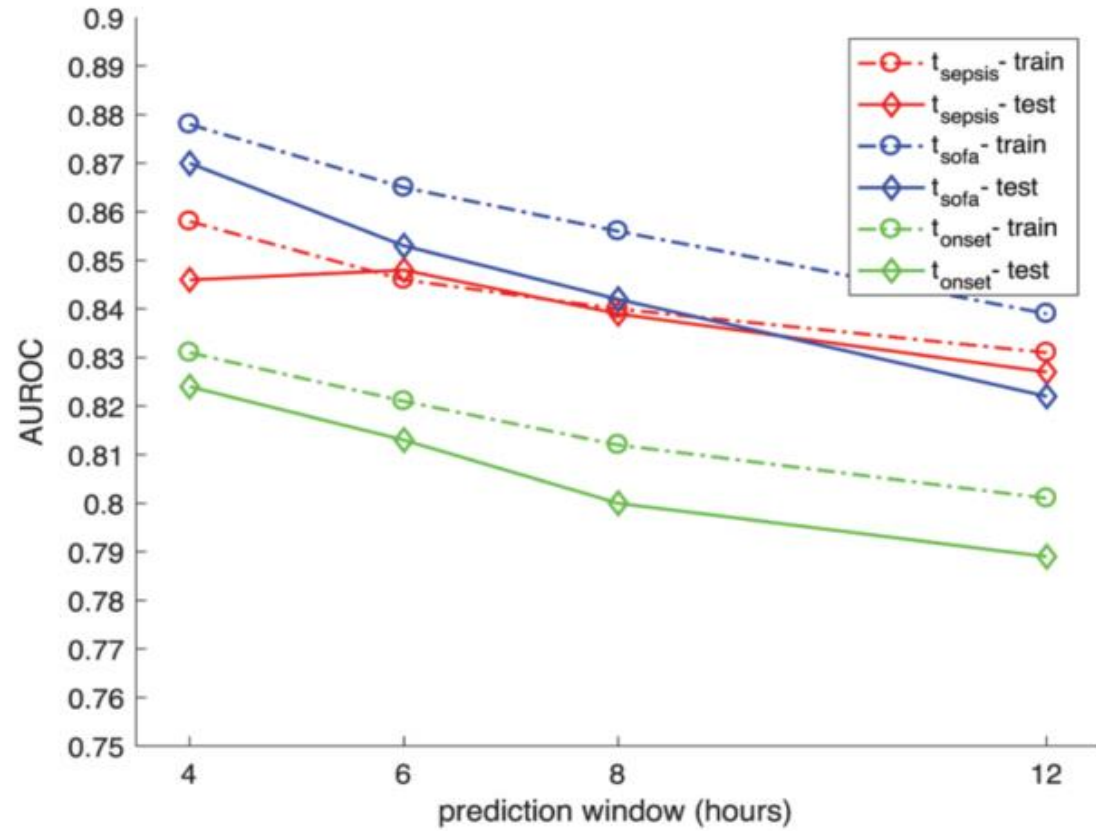
Relevance Score with Interpretability

1. 모든 특성 변수에 대한 Sepsis Risk Score의 편미분 계산 (gradient vector)
2. 특정한 특성 변수의 벡터와 해당 변수의 gradient vector 곱함
→ 각 입력 변수들이 **Sepsis Risk Score**예측에 얼마나 기여하는 지 계산 가능

$$f(x) = \sum_{i=1}^d R_i$$

- 총 d개의 특성 변수 존재
- $f(x)$ 라는 출력을 얻기 위해, 개별 특성 변수들이 기여하는 정도를 계산

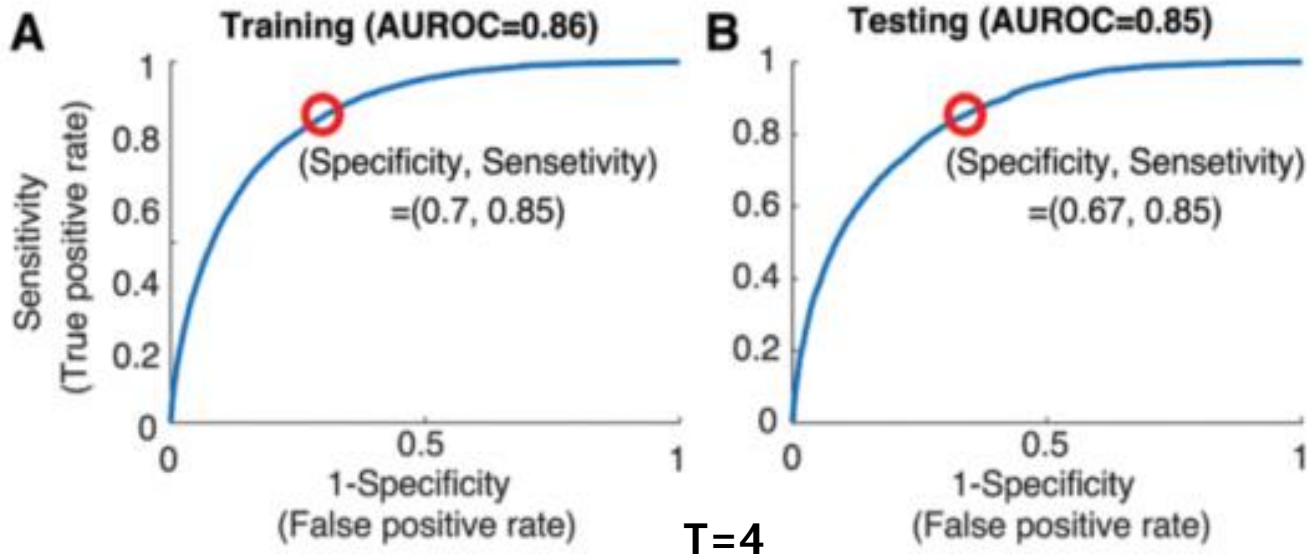
Results



* 실선 : test set / 점선 : train set

- 총 4가지 case의 prediction window ($T=12, 8, 6, 4$) 중 $T=4$ 인 경우의 예측 성능이 가장 우수 했음
- 대체적으로 과적합의 정도가 낮음 (일반화 가능성 높음)

Results



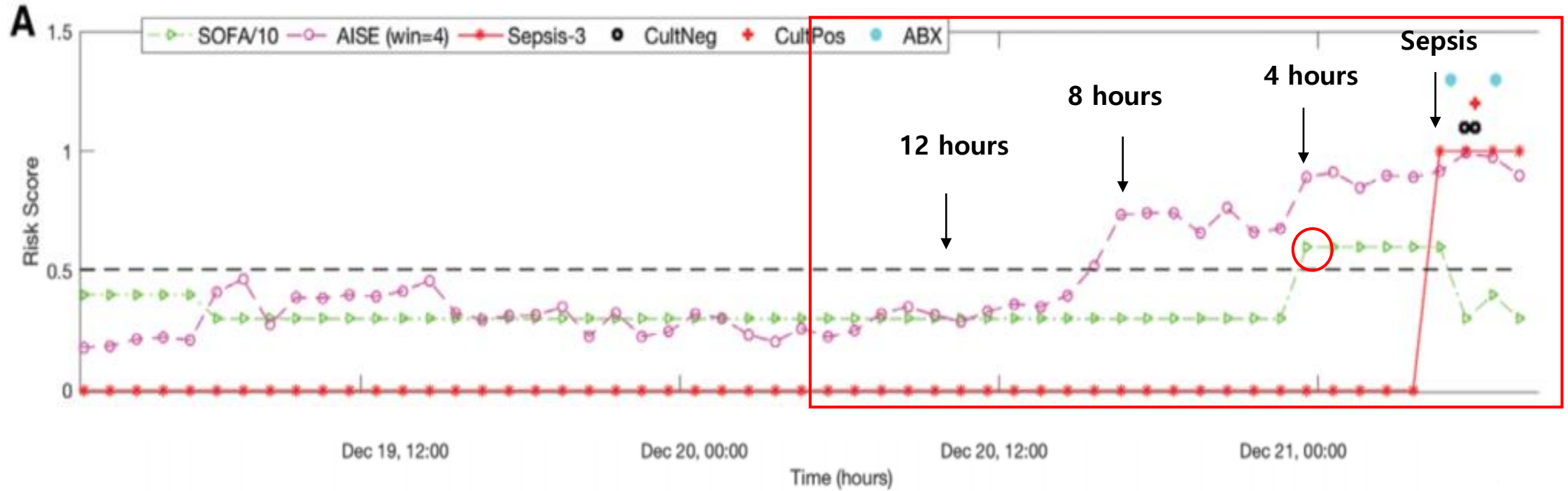
* 민감도를 0.85 수준으로 고정 (risk score=0.45)
(민감도와 특이도는 대체로 trade off 관계)

- **민감도** : 패혈증이 있는 사람을 얼마나 잘 찾아내는가 $TP / (TP+FN)$
- **특이도** : 정상인 경우를 얼마나 잘 찾아내는가 $TN / (TN+FP)$

TP	FN
FP	TN

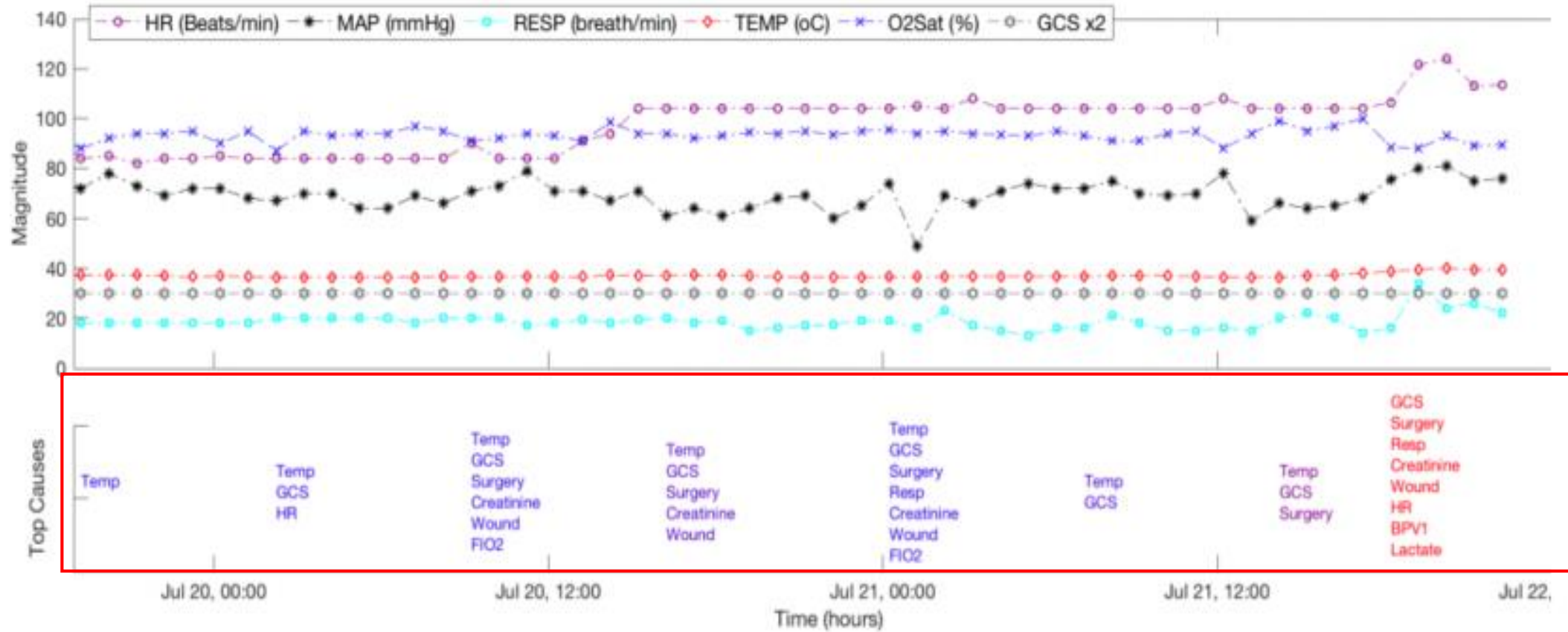
- A : 민감도 0.85 (패혈증 환자의 85%를 예측) / 특이도 0.7 (30%의 false alarms)
- B : 민감도 0.85 / 특이도 0.67 (33%의 false alarms)

Results



- 패혈증 발병 8시간, 4시간 전의 AISE 예측 성능이 우수
- 패혈증 발병 4시간 전의 SOFA 예측 성능이 우수 – p16 참고
- Prediction Window가 클 수록 예측 성능이 저하

Results



- 특정 시계열 point에서의 vital sign 측정치(위) AISE score에 기여도가 높았던 변수들(아래) 표시 → **Interpretability**

Results

1. 남성 환자가 여성 환자보다 패혈증 발병 가능성이 높음 (56.2% vs 54.2%)
2. CCI 지수가 높은 (4 vs 2) 환자의 패혈증 발병 가능성이 높음
3. ICU 체류기간 (5.9d vs 1.9d), SOFA의 중앙값 (5.0 vs 1.7), 병원 내 사망률(14.5% vs 2.9%)이 높음
4. Sepsis Risk Score가 높을 수록 사망 가능성이 높음

→ **EMR data와 high-resolution 시계열 data를 모두 활용했을 때,**

T=4인 case에서 AUROC가 0.85로 높은 분류 성능을 보이는 모델을 구축할 수 있음

Discussion

머신 러닝의 장점

1. 머신 러닝은 데이터에 내재되어 있는 패턴과 규칙을 활용하여 결과를 예측
2. **고빈도로 수집되는 real time data**가 증가함에 따라, 연구와 임상실습에서 머신 러닝의 중요도는 더욱 높아질 것
3. 머신 러닝은 특성 변수와 타겟 변수의 **복잡하고 비선형적인 관계**를 설명할 수 있는 강력한 툴을 제공
→ 전통적 통계 모형이 갖는 **과적합의 문제**를 효과적으로 해결

AISE 알고리즘의 장점

- 서로 다른 **resolution의 dataset**을 결합 (low resolution EMR + high resolution)
→ 패혈증 발병 예측의 학습에 있어 보완적인 정보 제공 (풍부한 정보량)
→ robustness of model
- 예측과 **예측의 근거** 제공
→ 시간의 경과에 따른 risk score에 **기여도와 관련성**이 높은 특징을 제공할 수 있음 - **Interpretability**

결론 및 의의

- ICU의 real time data를 활용 함으로서 AISE 알고리즘은 패혈증의 발병을 4-12시간 전에 높은 성능으로 예측할 수 있다
- 해당 알고리즘은 높은 Interpretability를 바탕으로 패혈증의 예측을 넘어선 예방의 영역으로 확장될 수 있다
 - 'Prediction is less important than interdiction' - **Interpretable AI**의 목적
 - 치료가 필요한 후보 환자들을 식별하여 발병률, 치료 비용을 완화할 수 있음
- 결함이 있는 data point(ex 사람의 수기로 작성)라도, **다른 측정치의 정보와 결합**된다면 모델에 사용되기에 적합

의문점

- Parameter 학습에 있어 Back Propagation 사용
→ 해당 모델이 Back Propagation을 사용할 만큼 복잡한 모델인지
- 패혈증 예측에 있어서 개별 변수들의 기여도 (relevance score)
→ 변수들간의 상호작용을 고려했는지, 고려했다면 상호작용의 정도와 방향을 설명할 수 있는지
- 시계열적 패턴, 추세가 고려되었는지
- 모델링, 모수 업데이트 알고리즘 보다는, Cohort Data의 구성에 초점을 맞추고 있음