

경영전문석사학위 논문

기계학습 모델을 이용한 중환자실
위양성 경보 절감에 관한 연구

2022년 8월

서울과학종합대학원대학교

현 중 균

경영전문석사학위 논문

기계학습 모델을 이용한 중환자실
위양성 경보 절감에 관한 연구

2022년 8월

서울과학종합대학원대학교

현 중 균

기계학습 모델을 이용한 중환자실
위양성 경보 절감에 관한 연구

지도교수 장 중 호

이 논문을 경영학 석사 학위논문으로 제출함

2022년 8월

서울과학종합대학원대학교

현중균

현중균의 석사 학위논문을 인준함

2022년 7월

위원장 신 호 상 (인)

위원 박 정 열 (인)

위원 장 중 호 (인)

초 록

본 연구는 의료진이 부족한 대한민국 의료환경, 특히 중환자실에서의 환자 감시장치의 높은 위양성 경보로 인한 의료진의 피로를 줄일 수 있는 의미 있는 경보지표 방법을 제안하였다.

현재 중환자실 환자감시장치는 주요 생체신호에 대해서 설정된 상한 값, 하한 값을 넘은 경우 경보를 울리게 되어있는데 많은 경보가 위양성이라는 선행연구에서 밝히고 있다. 이러한 위양성 경보에 대한 진위여부 판단을 돕기 위한 방법으로 생체신호 측정시점에서 3시간 뒤 환자의 위급상태(본 연구에서는 패혈증 쇼크 발병여부)와 비교하는 방법을 고려하였다.

연구에 사용한 데이터세트는 미국내 중환자실 데이터를 집대성한 MIMIC-3 데이터와 eICU 데이터를 병합하여 사용하였다. 쇼크예측에 사용한 인공지능 알고리즘은 XGBoost, Random Forest 및 Multi-layer Perceptron 을 사용하여 예측 인공지능 모델을 개발하였다.

경보에 대한 위양성 경보 검증 방법은 측정시점의 생체신호가 해당 생체신호에 설정된 경계 값을 넘어 경보가 울리는 조건이 되었을 때 3시간뒤 환자의 쇼크발생이 기록된 경우 True Positive 경보, 경보가 울리는 조건이었을 때 3시간뒤 환자의 쇼크발생이 기록되지 않았을 경우 위양성 경보로 정의하고 그 회수를 비교하였다.

총 5,698건의 데이터를 비교하였고 이중 1,799건이 설정된 경계 값을 넘

은 경보가 울리는 경우였고 이런 경우 3시간뒤 패혈증 쇼크가 기록되지 않은 건은 624건이었다. 이는 약 34.6%의 경감을 의미한다.

중환자실의 환경상 환자감시장치의 경고 범위 설정 범위나 운영상의 방법에 따라서 그 차이는 있을 수 있으나 본 연구의 결과는 측정시간이 가까운 미래의 환자의 상태와 비교하여 현재의 환자상태를 판단하는 의사결정 보조 지표로도 가능할 것이라는 가능성을 보았다고 할 수 있다.

본 연구에서 사용된 데이터세트는 백인위주의 데이터이고 한국의 중환자실 운영상황과는 상이할 수 있는 데이터세트로 국내 의료환경에서 수집된 데이터세트를 활용한 임상을 통해 국내 의료환경에서의 적용 가능성을 보인다면 의료진이 부족한 국내 의료환경에 조금이나마 도움이 될 수 있을 것이라 사료된다.

목 차

제 I 장 개요.....	1
제 1 절 연구배경.....	1
제 II 장 선행 연구.....	4
제 1 절 기존연구.....	4
제 2 절 기존연구 한계.....	6
제 3 절 연구 제안 요약.....	7
제 III 장 데이터 및 기법.....	9
제 1 절 연구 데이터.....	9
(1) 데이터 세트.....	9
(2) 데이터 처리.....	10
제 2 절 연구 기법.....	15
(1) 데이터세트 구조.....	15
(2) 비교방법.....	16
(3) 예측 모델 특징.....	17
제 III 장 결과.....	22
제 1 절 예측 결과.....	22
제 2 절 비교 결과.....	23
제 IV 장 토의 및 결론.....	24
제 1 절 결과 해석 및 의견.....	24
제 2 절 한계 및 미래 연구 제안.....	24

표 목차

<표 1> 입력변수 별 정상범위	13
<표 2> 데이터셋 구성(undersampling 전)	14
<표 3> 최종학습데이터 구성(undersampling 적용 후)	15
<표 4> 입력변수 별 경계값.....	16
<표 5> 알고리즘 별 패혈증 쇼크 예측 성능치 모음	22

그림 목차

<그림 1> Examples of Paired and unpaired vital sign Outliers(David M. Maslove,2016)	6
<그림 2> 데이터세트 추출 흐름도	12
<그림 3> 데이터세트 구조	14
<그림 4> 학습데이터세트 구조	15
<그림 5> XGBoost 알고리즘 개념도	18
<그림 6> Random Forest 알고리즘 개념도	18
<그림 7> MLP 알고리즘 개념도	19
<그림 8> 5 중첩 교차검증 개념도	20
<그림 9> ROC 및 AUC 도해	21
<그림 10> MIMIC-III 환자 인종 분포도 (Zheng Dai)	25
<그림 11> eICU 환자 인종 분포도(Tom J 외)	26

제 I 장 개요

제 1 절 연구배경

2021년 보건복지부가 발행한 OECD Health Statistics 2021에 따르면 대한민국은 2019년 기준 인구 1,000명당 임상 의사수는 2.5으로 OECD국가 중 폴란드, 멕시코에 이어 최하위에 달하는 반면 병상수는 12.8로 일본과 더불어 최상위에 위치하고 있다. 즉, 임상 의사당 관리해야 하는 병상수는 OECD국가에서 가장 높다. 더욱이 대한공중보건협회의에 따르면 2020년부터 팬데믹 영향으로 의료진의 부족은 더욱 심해졌다. (청년의사, 2021)

의료영역에서 가장 위중한 환자들을 관리하는 중환자실은 환자의 생명을 유지하고 주요 장기손상을 최소화하기 위해 24시간 고강도 집중치료가 시행되는 곳이며 환자 1인당 3.80+-2.80대 장비로 관리를 받고 있다(정유진, 김현정). ECRI(2015)는 2015년 이러한 의료장비에서 발생하는 경보음은 2014년 응급 치료 연구기관 보고서에서 가장 위험한 보건 의료 기술 위험요소로 선정하였다.

환자의 생명을 관리하는 장치는 관리대상환자의 수치가 일정범위를 넘어서면 경고를 울리는 기능을 필수적으로 가지고 있고 실제 현장에서 필수적으로 사용하고 있다. 그러나, 환자의 상태를 파악하는 이러한 장치에서 발생하는 경보 중 정확한 생리학적 위험을 나타내는 데이터가 아님에도 발생하는 경보음을 일컫는 위양성 경보음(False Alarm)은 현대의학을 활용하는 전 세계 응급 및 병원 현장에서 오랫동안 응급환자 대응에 심각한 문제가 되고 있는데 이는 아직도 미해결 상태로 남아 있다 (Borowski M 외 5, 2011).

이러한 위양성 경보음에 대한 문제는 국내만의 문제가 아닌 미국내 의료환경에서 병상당 1일 188건의 경보가 울리며 이중 88.8%가 위양성 경보라는 점을 들어 “경보로 인한 피로(Alarm fatigue)”가 환자에게 안정적인 환경을 만들어 주지 못한다고 밝혔다(Barbara J. Drew 외 1, 2014).

ECRI(2019)는 2014년에 이어 2019년에도 의료환경에서의 각종 경보에 대한 피로도(Alarm, Alert, and Notification Overload)를 Top 10 Health Technology Hazards for 2020에 포함시켰다.

다른 연구에서는 의료기기의 위양성 경보의 경우 병원 단위의 경보 중 80-99%가 거짓이거나 임상적으로 중요하지 않아, 의료진의 피로도를 높여 정말 중요한 경보에 대한 대응을 늦게 만든다고 하였다(Chrystinne Fernandes 외 2, 2020).

의료현장에서 경보음은 크게 타당한 경보음과 위양성 경보음으로 구분할 수 있다. 타당한 경보음의 이론적 정의는 환자의 상태 변화나 문제 상황에 대해 의료진에게 주의를 환기시켜 환자의 안전을 향상시키기 위한 경보음을 이야기한다(Healthcare Technology Foundation). 조작적 정의로서 타당한 경보음은 정확한 생리학적 데이터의 진정한 위반을 나타내는 경보음을 의미한다(Graham K 외 1, 2010). 반면 위양성 경보음의 이론적 정의는 환자의 상태 변화가 없는 상황에서 발생하는 경보음으로 임상적으로 즉각적인 중재가 필요하지 않는 경보음으로 정의한다(Park MY, 2016). 위양성 경보음의 조작적 정의는 분명한 생리학적 데이터의 위반 없이도 나타내는 경보음을 이야기한다(ECRI, 2007).

위양성 경보음은 비기술적 경보음과 기술적 경보음으로 나뉘어지는데 환자의 움직임 증가, 간호행위 및 시술과정에서의 측정 오류 시, 심전도 측정에서 잡음을 잘못 인식하는 상황 시, 환자의 말초 순환 장애로 산소포화도 정확도가 떨어져 발생하는 경보 등이 전자의 경우이며, 환자에게 부착된 전극 및 세서의 부착 불량으로 유발되는 상태가 후자의 경우이다.

이러한 위양성 경보음은 의료진의 경보음에 대한 경각심을 낮추는 결과를 낳게 하여 의료인의 즉각적인 처치가 필요한 심각한 상황에서도 일어날 수 있기에 환자안전에 큰 문제가 될 수 있다.

Eric K Wei와 2인이 자마네트워크(JAMA Network)에 기고한 사설(2021)에서 코로나 19가 병원에 준 9가지 교훈 중 하나로 ‘병원 출입 최소화하면서도 환자 상태 확인할 수 있어야 한다’ 라는 교훈으로 환자들의 상태를 병실에 잤은 출입 없이도 확인할 수 있도록 병실 구조에 변화를 줘야 하며 이를 위해 병실 벽에 창문으로 설치하거나 문을 유리문으로 교체, 비디오 모니터링을 활용하는 방법 등을 제안했다. 이는 위양성 경보로 인해 의료진이 환자상태를 확인하려 병실에 들어가야하는 횟수를 줄이도록 경보의 진위성을 판정할 수 있는 방안은 이 교훈에 일맥상통한다 할 수 있겠다.

본 연구에서는 상기에서 언급한 의료진의 부족과 코로나로 인한 의료진과 환자의 불필요한 접촉을 최소화하기 위한 방안으로 환자감시장치에서 발생하는 생체신호 중 일부를 기계학습을 이용하여 환자의 미래상태(쇼크 여부) 예측값을 산출하고 이를 경보음 시점과 비교하여 경보음의 진위성을 판단할 수 있는 새로운 방법을 고안하였다.

제 II 장 선행 연구

제 1 절 기존연구

그 동안 의료현장에서는 중환자실에서 사용되는 의료장비의 위양성 경보에 대한 실체를 연구하고 그 대안점을 연구하였다.

정유진 외 1(2018)은 그들의 논문에서 조사한 자료에 의하면, 의료 현장에서의 '경보음 발생 빈도 및 위양성 경보음 비율'은 타당한 경보음 40.2%, 위양성 경보음 59.8%로 나타나, 무려 60%가 의미 없는 경보음이라는 통계 수치가 집계됨으로써 그 심각성을 확인할 수 있었다.

의미 없는 경보음의 환경에서 의료진들은 경보음에 진지하게 대처하지 않게 되도록 만들어지고 있다.

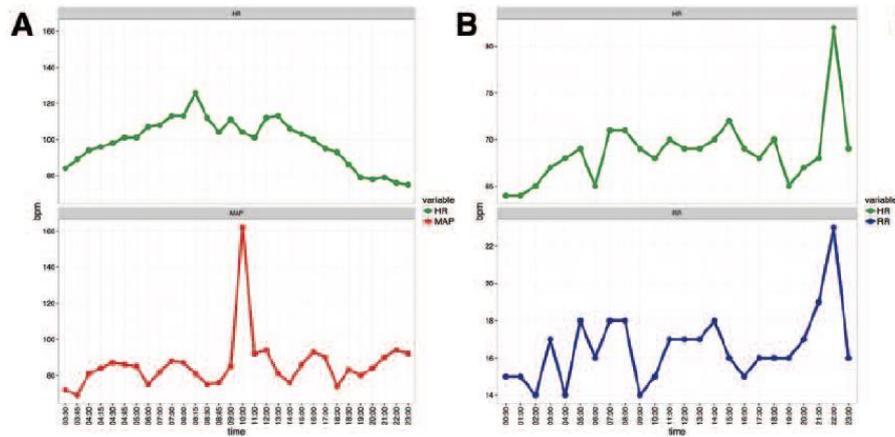
정유진 외 1(2018)의 논문에서는 타당한 경보음이 울린 경우 무반응이 43.2% (32건), 경보음의 원인 해결이 27% (20건), 2분 경보음 음소거 24.3% (18건)이었고 위양성 경보음인 경우 경보음 관리로 2분 경보음 음소거가 47% (52건), 무반응 25.5% (28건), 경보음 원인 해결 24.5% (27건) 이었다고 밝혔다.

응급 및 병원 현장의 구조사 혹은 간호사들은 여러 모니터링 기기를 관리하고 경보음에 반응하는 핵심 인력들로서 의료장비 경보음에 의해 가장 직접적 영향을 받는 사람들이지만, 이들이 접하는 수많은 의미 없는 경보음에 쉽게 익숙해져서 경보음 대부분이 거짓이거나 조치를 하지 않아도 된다고 여기게 되어 경보음에 즉시 대응하지 않는 경향이 있다. 따라서 이들은 경보음을

비활성화하거나, 경보음의 범위를 너무 넓게 설정함으로써 일부 중대하고 생명을 위협하는 상황(예: 쇼크 상황)에서 경보음이 감지되지 않는 문제가 발생할 수 있으며 이로 인해 환자안전에 대한 심각한 위협이 제기되고 있으며 수많은 기관들은 환자안전에 중대한 문제로 위양성 경보음 (False Alarm)의 위험을 인식하고 있는 실정이다. (Ok Min Cho,2016)

Borowski M 외 5인(2011) 등도 정확한 생리학적 위험을 나타내는 데이터가 아님에도 발생하는 경보음을 일컫는 위양성 경보음은 현대의학을 활용하는 전 세계 응급 및 병원 현장에서 오랫동안 응급환자 대응에 심각한 문제가 되고 있다고 기술했다.

David M. Maslove 외 3인(2016) 는 앞서의 논문과 달리 환자감시장치의 생체신호 간의 상관관계에 대한 고찰로 경보의 진위성과 위양성을 파악하고자 하는 연구가 있었다. 예를 들어 심박수와 평균동맥압(MAP, mean arterial pressure)의 측정값 중 하나가 상당한 측정값의 차이를 보이는 경우 이상치 또는 위양성 경보로 구별하거나 호흡수와 심박수가 급격히 같이 오르는 경우 인 경우는 그 값을 평균적인 추세에 비해 급격히 늘었으나 위양성 경보로 보지 않은 경우 등이다.



<그림 1> Examples of Paired and unpaired vital sign Outliers (David M. Maslove, 2016)

제 2 절 기존연구 한계

기존 연구에서는 중환자실 환경에서의 위양성 경보에 대한 통계수준으로 현황을 파악하기 위한 것이고 다른 연구는 동일시간대 입력 변수 간의 상관관계가 크게 벗어나는 경우를 고려한 것이 주된 연구였다.

미국의 TigerConnect사는 경보에 대한 피로를 줄이는 방안으로 다음과 같이 제시하였다.

- 알람 피로의 위험에 대해 직원 교육
- 보안 메시징 시스템을 사용하여 알람을 적시에 정확하게 라우팅
- 경보관리 소프트웨어와 미들웨어 플랫폼을 동시에 활용한 통합 솔루션 도입
- 알람 소리가 다른 우선 순위 또는 심각도에 따라 구별되도록 보장

- 다양한 심각도 경보를 적절한 수준 및 유형으로 구분하여 직원에게 전달

이들이 개발한 내용은 의료 팀 협업 플랫폼 형태로 중앙 집중식 데이터 및 워크 플로를 실시간 환자 정보와 결합하여 사용하는 방식으로 구성되는 등의 프로세스 개선 방향 솔루션이다.

지금까지의 선행연구 및 개발된 솔루션은 현재의 측정된 신호가 경계치를 벗어난 경보에 대한 통계 및 해결책이 주를 이뤘다.

무엇보다 기존 환자감시 장치에서 얻어지는 실시간 데이터로는 지금의 문제 즉 현재 주어지는 경고음의 의미만 있을 뿐 가까운 미래에 대한 환자상태를 보여주는 것은 아니다. 본 연구의 핵심가설은 ‘만약 지금의 환자감시장치에서 울리는 경보가 환자의 가까운 미래 상태와의 상관관계를 정량화 할 수 있고 이에 따른 위양성 경보판정에 신뢰적인 지표를 제공한다면 의료진의 피로도를 줄이는 방안이 될 수 있을 것’이며 이를 기계학습 방법을 통해 증명하고자 하는 것이 본 연구의 목적이다.

제 3 절 연구 제안 요약

본 연구에서는 중환자실 입실한 환자에게 연결된 환자감시장치의 주요 생체신호 데이터(수축기혈압(SBP), 이완기혈압(DBP), 호흡수(RR), 심장박동수(HR), 체온(BT), 산소포화도(SpO2))를 기반으로 환자의 위급상태를 예측하는 알고리즘을 개발하고 이를 바탕으로 경보가 울리는 현재의 상태와 가까운

미래(본 연구에서는 3시간)의 환자 상태 예측 값과 비교하여 특정시점에서 발생한 경보에 대한 의료진의 판단을 위해 보조적 지표로 활용 가능할 수 있는지 판단하고자 하였다.

본 연구를 위한 환자데이터세트는 중환자실 내에는 패혈증 판명 환자로 제한하였다. 패혈증은 장기에 염증이 발생하는 질병으로 전신으로 발생되면 다발성 장기부전이 발생하게 되고 이로 인해 패혈성 쇼크가 발생하게 된다. 패혈증 쇼크는 감염에 혈류역학적 불안정성 상태를 말하며 패혈증 환자의 70% 이상은 중증 패혈증 또는 패혈성 쇼크로 진행된다. (이주리 외 1, 2014)

중증 패혈증은 미국에서 비심장계 중환자실 사망률의 30~50%를 차지하며, 전체 사망의 주요 원인이며, 우리나라도 중증 패혈증으로 인한 사망률이 30%에 이르는 등 패혈증과 패혈성 쇼크는 전 세계적으로 중환자에게 가장 높은 사망률을 나타내는 원인이다. (나선경 외 1, 2013)

본 연구에서는 중환자실에서 환자의 상태에 많은 비중을 차지하는 패혈증 환자의 데이터로 제한두고 수행하였다.

본 모델을 얻기 위하여 의료데이터연구에서 자주 등장하는 MIMIC-III 데이터 및 eICU 중환자실 데이터를 이용하였고 기계학습 알고리즘 2가지와 심도학습의 알고리즘 1가지를 수행하였다.

제 III 장 데이터 및 기법

제 1 절 연구 데이터

(1) 데이터 세트

본 연구를 위해서 다음 2종류의 데이터 세트를 혼합하여 사용하였다.

1) MIMIC-III 데이터 세트

MIMIC-III는 Medical Information Mart of Intensive Care-3의 약자로 Beth Israel Deaconess Medical Center 중환자실에서 발생한 데이터를 MIT에서 정리하고 개발된 단일 센터 데이터베이스이다. 2001~2012년 사이에 약 4만 명 이상의 비식별화 된 보건의료 데이터를 포함하며 인구통계자료(demographics), 생체 신호, 실험실 결과, 처방전, 간병인 메모, 이미지 기록, 사망 여부 등의 총 26개의 테이블로 구성되어 있다.

2) eICU 데이터 세트

eICU 데이터세트는 Collaborative Research Database의 약자로, Philips Healthcare와 MIT Laboratory for Computational Physiology가 공동으로 2014~2015년 사이에 미국 전역의 ICU에 약 200,000건 이상의 입원과 관련된 비식별화 된 보건의료 데이터로 구성된 다중 센터 데이터베이스이다.

이 데이터베이스는 생체 신호, 치료 계획 문서(care plan documentation), 질병 측정의 심각도(severity of illness measures), 진단 정보(diagnosis

information) 및 치료 정보(treatment information) 등의 데이터로 구성되어 있으며, eICU 데이터베이스도 연차별로 갱신되고 있다. 본 연구에서는 2018년 발표된 데이터를 사용하였다.

(2) 데이터 처리

1) 데이터 선별 기준

본 연구의 학습데이터를 구성하기 위해서는 다음과 같은 기준으로 1차 선별을 수행했다.

a. 18세 이상의 성인환자

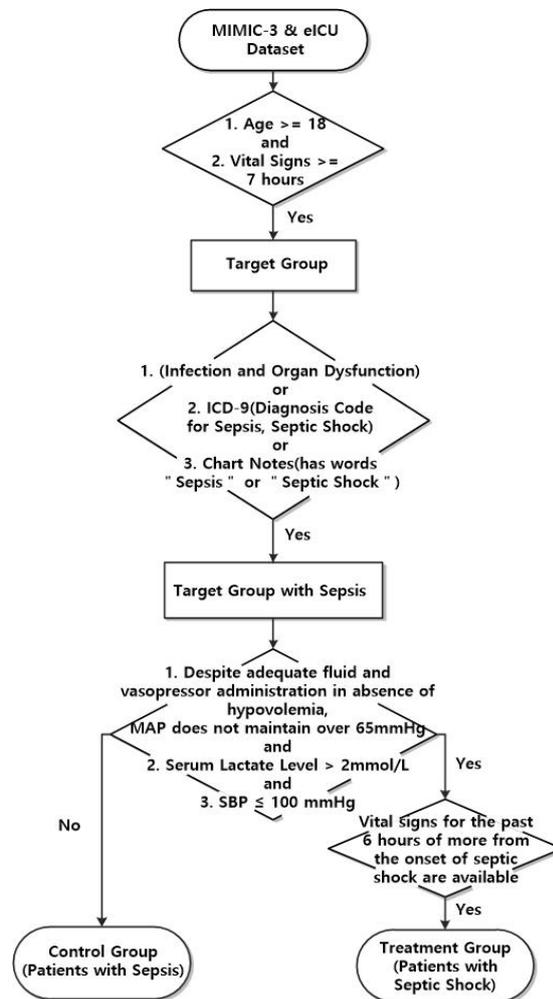
b. 중환자실 입실 후 7시간 이상의 생체신호가 기록된 환자 데이터

7시간 이상의 생체신호가 기록된 환자데이터로 제한한 것은 3시간 뒤 환자상태를 예측하기 위하여 측정시간 이전 3시간 분량의 데이터가 필요하기 때문이다.

상기 환자 데이터세트 내에서 3시간 뒤 환자의 패혈증 쇼크의 여부를 결정하는 기준으로는 ICD-9 코드와 Angus ICD 9-CM에서 패혈증관련 질병코드가 입력되어있는 환자와 차트노트에 sepsis, severe sepsis, 패혈증 쇼크가 기록된 환자의 레코드만을 수집하여 2차 선별을 수행했다.

2차 선별된 데이터세트에서 혈량저하증(Hypovolemia)가 없는 상태에서 수액과 바소프레신(Vasopressin) 계열 승압제를 투입했지만 MAP가 65 미만

이고 혈청내 젓산 수치가 2mmol/L 이상이면서 SBP(수축기 혈압)가 100mmHg 이하인 환자의 데이터는 패혈증 쇼크 환자군(실험군, Treatment Group), 그 반대되는 환자군(대조군, Control Group, 패혈증 쇼크가 발생하지 않은 환자)로 라벨링을 하여 <그림 2>와 같이 학습데이터세트를 구성하였다. 여기서 바소프레신 계열의 승압제는 뇌하수체 뒤엽 호르몬의 하나로 콩팥에서 수분의 재흡수를 촉진하고 모세 혈관을 수축시켜 혈압을 높이는 작용을 한다. (바소프레신, 위키피디아, 2022)



<그림 2> 데이터세트 추출 흐름도

2) 이상치

본 연구에서 사용된 상기 2종류의 데이터베이스는 수기 입력방식으로 입력시 오류와 기기의 오류로 범위를 벗어난 데이터를 포함하고 있다. 이상치를 찾는 통계 기법 중의 하나로 Robust Z-Score(이하 “Z-Score”)를 사용하여 정상 범주에서 지나치게 벗어나 있는 데이터를 찾아냈다. 정상 범주의 범위는

Z-Score가 3.5 이상인 값에 대해서 null처리 후 선형 보간법을 진행하여 값을 수정하였다. Z-Score 3.5 이내의 입력변수 범위는 다음의 표와 같다.

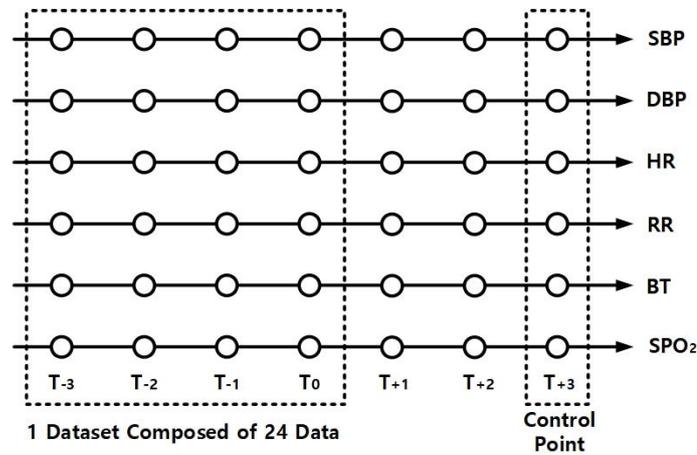
입력변수 별 정상범위는 다음의 표1 과 같다.

<표 1> 입력변수 별 정상범위

Parameter	입력 범위
SBP	34 < SBP < 300
DBP	12 < DBP < 200
HR	23 < HR < 250
RR	5 < RR < 80
Tempc	25 < Tempc < 45
SpO ₂	60 < SpO ₂ < 100

3) 결측치 처리 방식

실제 중환자실 환경에서도 입력이 누락되는 경우가 발생한다. 훈련데이터 세트에서 결측치는 carry forward 방식으로 최근 데이터를 그대로 가져오는 방식을 취하여 가장 쉬운 결측치 방식을 선택하였다. 결측치를 채워야하는 이유는 본 연구는 측정시간에서 3시간 이후 패혈증 쇼크가 발생할 확률을 예측하기 때문에, 1시간 간격으로 현재부터(T0) 과거 3개 시간 데이터(T-1~T-3)를 사용하여 학습시켜야 한다. 그래서 기준점(T+3)으로 지정할 수 있는 생체 신호와 마지막 시점인 생체 신호(T-3)까지에 대한 총 7시간 이상의 생체 신호가 필요하다 (그림 2).



<그림 3> 데이터세트 구조

4) 최종 데이터세트

지금까지의 추출 과정을 거쳐 실험군과 대조군의 데이터 수는 <표 2>와 같다.

<표 2> 데이터세트 구성 (undersampling 전)

구분	실험군 (패혈증 쇼크가 발생한 패혈증 환자군)	대조군 (쇼크발생하지 않은 패혈증환자군)	합계
MIMIC-3	1,355	3,387	4,743
eICU	1,494	4,482	5,976
합계	2,849	7,869	10,719

여기서 대조군의 데이터수가 실험군에 비해 2.76배정도 더 많은 불균형 데이터 세트에 실험군의 데이터 수에 맞게 대조군을 undersampling 하여 데이터세트를 5:5로 구성하였다. Undersampling의 이유는 불균형한 데이터세트로 학습을 수행하면 가중치가 높은 클래스를 더 예측하기 때문에 정확도 (accuracy)는 높아지지만 가중치가 낮은 클래스에 대해서는 정밀도

(precision)은 낮아질 수 있다.

<표 3> 최종학습데이터 구성 (undersampling 적용 후)

구분	실험군 (패혈증 쇼크가 발생한 패혈증 환자군)	대조군 (쇼크발생하지 않은 패혈증환자군)	합계
합계	2,849	2,849	5,698

학습데이터세트와 검증데이터세트를 각각 80%:20%로 구성하여 기계학습 데이터로 사용하였다.

제 2 절 연구 기법

(1) 데이터세트 구조

본 논문에서 패혈증 쇼크 발생 확률 예측을 계산하기 위하여 사용한 학습 데이터세트는 다음과 같은 형태로 이뤄졌다.

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
icustay_id	label	sbp3	sbp2	sbp1	sbp0	dbp3	dbp2	dbp1	dbp0	mbp3	mbp2	mbp1	mbp0	hr3	hr2	hr1	hr0	rr3	rr2	rr1	rr0	tempc3	tempc2	tempc1	tempc0	spo23	spo22	spo21	spo20
283746	1	101	114	104	104	56	61	54	54	71	78.7	70.7	70.7	72	69	70	74	13	21	20	18	36.3	36.5	36.3	36.2	90	91	90	89
283159	1	120	97	111	115	33	42	29	29	62	60.3	56.3	57.7	62	60	68	66	17	16.5	16	16.5	37.1	37.1	37.2	37.3	98	98	100	98
208154	0	125	125	125	125	67	67	67	67	86.3	86.3	86.3	86.3	99	99	99	99	22	22	22	22	36.9	36.9	36.9	36.9	93	93	93	93
270086	0	87	93	106	111	50	48	51	57	62.3	63	69.3	75	77	77	84	85	20	18	25	17	36	35.7	35.9	36.1	97	97	98	98
203378	0	117	117	118	119	60	61	62	63	78.9	79.8	80.7	81.6	112	112	121	118	24	28	26	24	36.6	36.6	36.6	36.7	96	99	100	100
207362	0	142	122	122	162	71	62	65	88	94.7	82	84	113	79	76	80	84	16	17	16	16	37.4	37.4	37.4	37.4	100	99	100	100
281750	0	155	180	183	185	78	88	89	90	104	119	120	122	92	93	93	93	20	19	19	19	36.8	36.5	36.4	36.2	87	90	91.5	93

<그림 4> 학습데이터세트 구조

현재시간의 수축기혈압(SBP)는 sbp0, 현재시간 기준 1시간전 수축기 혈

압은 sbp1, 두시간 전 수축기 혈압은 sbp2, 3시간 전 수축기 혈압은 sbp3로 표현하였다. 이완기 혈압(dbp), 호흡수(rr), 심박수(hh), 체온(tempc), 산소포화도(spo2)도 동일한 방식으로 표시하였다. Icustay_id는 중환자실 입원시 제공되는 id로 한 환자가 2번 이상 입실하는 경우 별도의 icustay_id가 발급된다.

Label열은 현재시간 기준 3시간 뒤 패혈증 쇼크 발생여부의 값으로 1은 쇼크 발생, 0은 쇼크 미 발생 경우로 정의했다.

입력 변수별 경계값은 표 4와 같고 이는 경보음이 울리는 기준으로 삼는다.

<표 4> 입력변수 별 경계값

측정대상	경계값
심박수(상한)	185
심박수(하한)	40
호흡수(상한)	40
호흡수(하한)	2
수축기혈압(상한)	195
수축기혈압(하한)	30
확장기혈압(상한)	180
확장기혈압(하한)	45
산소포화도(상한)	100
산소포화도(하한)	87
체온(상한)	35.0
체온(하한)	38.5

(2) 비교방법

비교 방법은 <그림 4>에서 sbp0, dbp0, rr0, hr0, spo20, tempc0의 값이 앞서 제시한 <표 4>의 경계값을 벗어났을 경우 해당 환자의 label의 값이 1

인 경우는 걱정 올림, 0인 경우는 위양성 경보로 정의하고 각각 비교하였다. 즉, 측정시점에서는 특정입력변수가 경계치를 넘어서 경보가 울렸지만 이 환자의 3시간 뒤 상태가 쇼크상태가 아니라면(label=0) 위양성 경보로 정의하고 그 횟수를 측정하는 것이다.

비교를 위해 <표 3>의 데이터를 데이터베이스(mysql)에 적재하고 해당 경계값을 위반하는 환자를 선별하는 SQL을 작성하여 그 수를 측정하였다.

(3) 예측 모델 특징

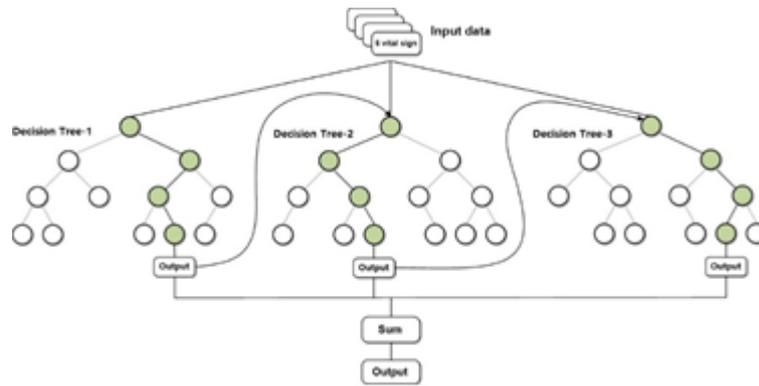
본 연구의 패혈증 쇼크 예측 모델은 <그림 3>에서 현재(T_0)부터 과거 3시간($T-3$)의 6가지 입력값을 데이터셋으로 구성하여 머신러닝 및 딥러닝 알고리즘으로 학습하여 3시간($T+3$) 뒤 패혈증 쇼크발생 예측확률에 대한 모델을 도출하였다.

본 연구에서 예측모델에 사용된 기계학습, 심화학습 알고리즘은 3시간 이후 패혈증 쇼크발생 확률을 구하는 모델을 개발하는데 사용되었다.

1) XGBoost

XGBoost는 다수의 결정 트리가 부스팅방식의 구조를 가지는 앙상블 모델이다. 앙상블 모델 내의 모든 결정 트리는 순차적으로 학습 및 예측이 진행하며, 모든 결정 트리는 이전 결정 트리의 잔차에 가중치를 부여하여 학습을 진행하는 알고리즘이다. 랜덤 포레스트의 예측 결과는 투표를 통해 채택하는 대

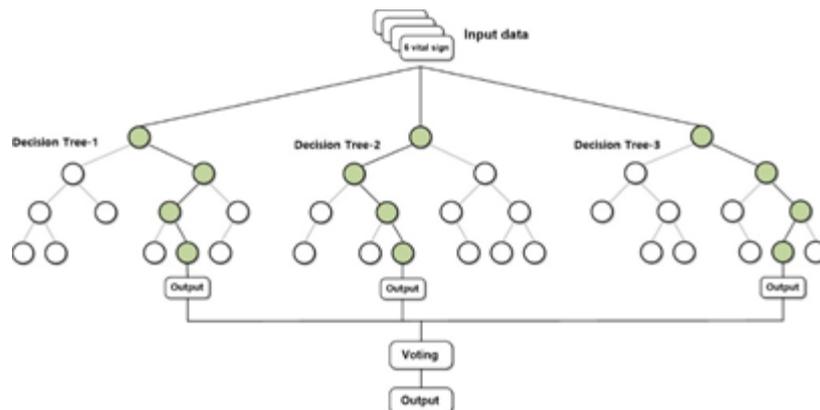
신, XGBoost의 예측 결과는 모든 결정 트리의 결과의 합계로 구해진다.
 (Chen 외 3, 2016)



<그림 5> XGBoost 알고리즘 개념도

2) Random Forest

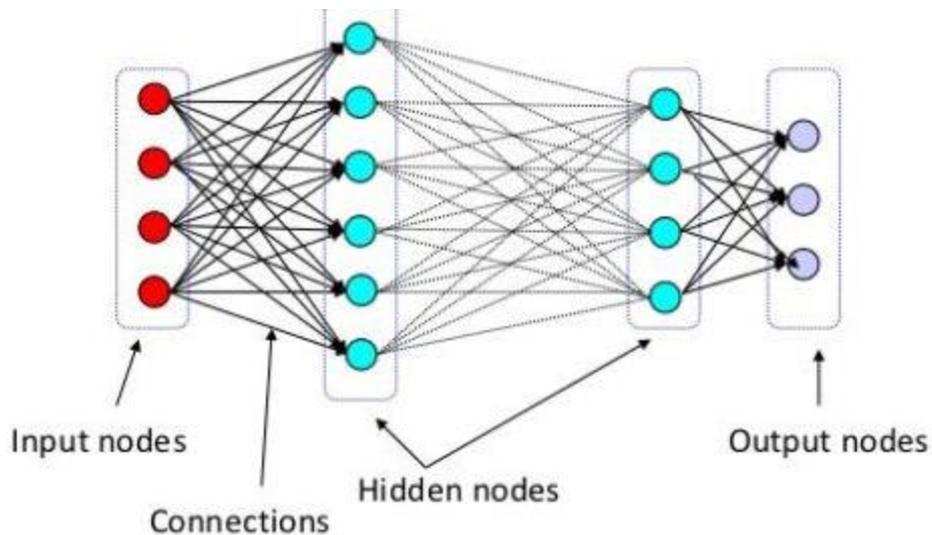
다수의 결정 트리가 배깅 (bagging) 구조를 가지는 앙상블 모델이다. 다수의 결정 트리가 입력 데이터에 대해 각각의 결과값을 도출하면, 투표(voting)을 진행하여 최종 결과값을 도출하는 알고리즘이다. (Breiman 외 1, 2001)



<그림 6> Random Forest 알고리즘 개념도

3) Multi-Layer Perceptron

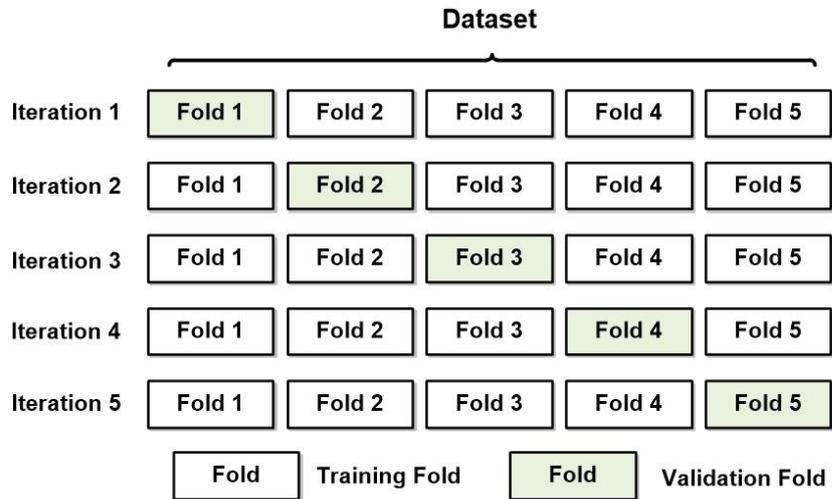
입력층(input layer)와 출력층(output layer)사이에 여러 개의 은닉층(hidden layer)들로 이루어진 네트워크이다. 시냅스의 결합으로 네트워크를 형성한 인공 뉴런(perceptron)이 학습을 통해 시냅스의 결합 세기를 변화시켜, 문제 해결 능력을 보여주는 알고리즘이다. (Schmidhuber 외 1, 2015)



<그림 7> MLP 알고리즘 개념도

4) 검증

학습 및 검증 과정에서 AI Model이 특정 데이터에만 편중되는 것을 방지하고, 일반화된 AI Model을 만들기 위해 k-fold 교차검증을 사용하였다. MIMIC-3 & eICU 데이터세트에서 라벨링된 데이터는 5-fold 교차검증을 사용하여 학습 및 검증 데이터세트로 구성하였다.

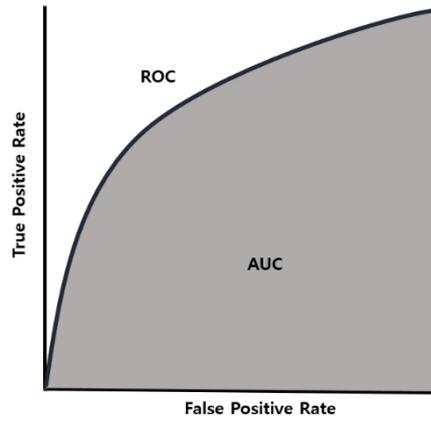


<그림 8> 5 중첩 교차검증 개념도

5) 평가 지표

기계학습 분류 모델에 대한 평가지표는 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 스코어, ROC 및 AUC 등이 있다. 이중 본 연구에서는 평가지표 AUC를 기준으로 예측도를 평가하기로 하였다.

AUC는 Area Under the ROC Curve의 약자로 ROC(Receiver Operating Characteristics) 아래의 면적을 뜻한다. ROC 커브의 X축은 양성이라 잘못 판정한 비율(False Positive Rate), Y축은 양성으로 정상 판정한 비율(True Positive Rate)으로 한 그래프를 의미한다. ROC커브는 분류기 성능을 2차원적 표현으로 분류기를 비교하고자 할 때 단일 스칼라 값으로 표현하는 것 중 하나가 AUC이다.



<그림 9> ROC 및 AUC 도해

AUC는 고정수치로 분류기 성능을 최소 0, 최대 1로 정량적으로 측정할 수 있는 장점이 있다.

제 III 장 결과

본 연구에서 목적하는 결과는 위양성 경보에 대한 비교이지만 이를 위해 예측 정확도가 필요하기 때문에 예측 모델 결과와 연구에서 목적하는 경보시 측정값과 예측값 간의 비교 결과를 기술하였다.

제 1 절 예측 결과

3가지 알고리즘 중 XGBoost와 Random Forest의 AUC가 0.86으로 MLP(0.80)보다 나은 성능을 보였다.

<표 5> 알고리즘 별 패혈증 쇼크 예측 성능치 모음

알고리즘	구분	Precision	Recall	Accuracy	AUC
XGBoost	대조군	0.77	0.78	0.77	0.86
	실험군	0.78	0.76		
Random Forest	대조군	0.79	0.78	0.79	0.86
	실험군	0.78	0.80		
MLP	대조군	0.74	0.71	0.73	0.80
	실험군	0.72	0.74		

각 알고리즘 별 대조군, 실험군에 대한 Precision과 Recall 값도 최대 3% 이하의 차이를 보였다.

제 2 절 비교 결과

주어진 데이터셋 총 5,658에서 측정시간(T0)의 <표4>에서 기술한 입력 변수의 최대, 최소 한계 값을 넘은 경우는 총 1,799건이며 이런 경우 3시간 뒤 쇼크 발병을 의미하는 label 값이 1인 경우는 총 1,175건이고 경보 대비 label이 0인 경우는 총 624건으로 집계되었다. 즉, 1,175건은 유의미한 경보일 수 있으며 34.6%의 624건은 무의미한 경보일 수 있다는 것으로 해석할 수 있다. 여기에 예측정확도 86%를 적용한다면 최소 536건에서 최대 624건 정도가 무의미한 경보로 여겨질 수 있다.

제 IV 장 토의 및 결론

제 1 절 결과 해석 및 의견

폐혈증 쇼크를 예측에 대한 성능지표 AUC는 XGBoost와 Random Forest 알고리즘이 86%로서 MLP의 80%보다 좋은 성능을 보였다. 또한 86%는 Mohaimenul(2018)에 비교된 8개 논문의 성능치와 비교했을 때 Q. Mao 외 9(0.92), H.J Kam 외 1(0.92), T. Desautels 외 10 (0.88) 다음의 순서였다.

본 연구에서 보여준 결과는 예측 값 대비 현재 측정되는 경보중 34.6%가 위양성 경보로 볼 수 있다고 확인되었다. 예측 성능값이 86%이라면 100건 중 16%는 위양성 예측으로 볼 수 있기 때문에 위양성 경보와 3시간뒤 예측 값 간의 차이에 86%를 곱한 값으로 오차에 대한 영향을 반영하였다.

더 많은 데이터를 기반으로 더 많은 알고리즘을 적용하여 성능을 높인 모델을 개발하여 위양성 경보를 줄일 수 있도록 또 다른 시도가 필요하겠다.

제 2 절 한계 및 미래 연구 제안

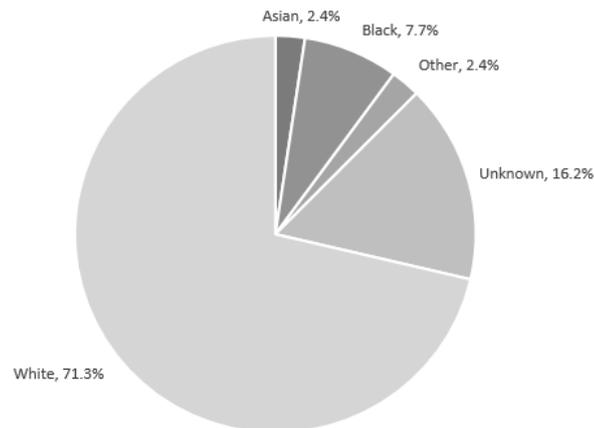
본 연구의 한계는 기계학습에 사용된 데이터셋을 구성하기 위해서 질병 코드, 차트 노트 등에 기록된 폐혈증 쇼크를 기준으로 선별 기준으로 사용하였다. 하지만 폐혈증 쇼크 기록시간은 실제 쇼크가 발생한 시간과 차이가 있을 수 있다. 폐혈증 쇼크는 MAP의 값이 65미만으로 떨어지는 경우로 볼 때 이 MAP가 65 미만으로 떨어지는 시간과 이를 차트에 기록하는 시간과는 큰 차이가 나지 않지만 의료진이 폐혈증 쇼크 기록을 기록하는 시점은 그 차이가

발생할 수 있다. 금번 연구에서 사용한 질병코드 차트기록만을 이용한 쇼크 발생 시점 보다 실제 의료현장에서 정확하게 기록될 수 있는 의미있는 기준값을 찾아 예측 모델링을 다시 수행해야할 것으로 보인다.

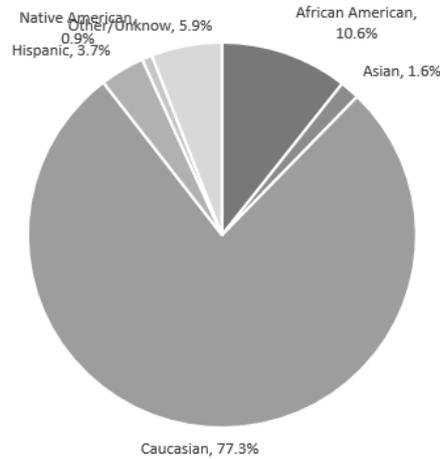
모델 구현을 위해 사용한 학습 데이터의 정상 범위를 조정하거나, 이상치에 대한 방법론을 적용해가며 모델에 대한 성능비교도 추가적으로 진행해야 할 사항이다.

한편, 학습에 사용된 두 데이터세트의 결측치에 대한 Carry forward 방법 한가지만 사용하여 수행되었고 다른 방법에 따른 예측 성능 비교가 없었다. 차후 연구에서는 여타 방법을 통해 성능 비교를 고려해야겠다.

또한 본 연구에서 사용된 두 의료데이터인 MIMIC-III와 eICU데이터는 백인 중심의 데이터 분포를 가지고 있어 각 국가별 환자, 특히 동양인이 절대적인 아시아 지역에서 쇼크예측 모델을 외부 검증이 필요하다.



<그림 10> MIMIC-III 환자 인종 분포도 (Zheng Dai)



<그림 11> eICU 환자 인종 분포도(Tom J 외)

이는 대한민국 중환자실 데이터를 이용하여 본 쇼크예측 모델에 대한 검증을 통해 범용성을 가진 모델인지 또는 지역적, 인종별 편향성을 가지는 과적합 모델이 될 수 있는지 검증 또한 필요할 것으로 판단된다.

본 연구는 위양성 경보에 대한 피로도를 낮추기 위한 하나의 방법론으로 제시한 것이다. 선행연구에서 제시했던 방법과는 다른 예측을 기반으로 하였다는 점에서 새롭다 할 수 있다. 이러한 방법이 실제로 의료현장에서 사용하기 위해서는 의료기기 허가를 득해야 한다.

소프트웨어 의료기기는 다른 산업의 솔루션 보다 현장도입을 위해서는 여러 임상단계를 거쳐야한다. 환자 상태의 예측치를 기준으로 현재 울리는 경보

음에 대한 진위 여부를 판단하는 지표 사용으로 보다 더 정확한 예측도와 전향적 임상 등을 통해 신뢰도를 가졌을 때에 의료기기 승인이 가능하며 실제 의료현장에서 사용가능한 추가적인 지표로 자리매김이 가능할 것으로 보인다.

끝으로 본 제안 방식의 위양성 경보 경감 기법은 현재 재직 중인 회사에서 개발 중인 감염병관련 소프트웨어 제품군인 qSOFA, SOFA 및 예측 솔루션에 이어 임상 경보관리 솔루션으로 자리매김할 수 있을 것으로 사료된다.

참고문헌

보건복지부 2021, *OECD Health Statistics 2021*, 한국보건사회연구원

ECRI Institute (2007), *The hazard of alarm overload: Keeping excessive physiologic monitoring alarms from impeding care. Health Devices.*

ECRI Institute (2014), *Top ten health technology hazards for 2015*

ECRI Institute (2019), *Top 10 Health Technology Hazards for 2020.*

Healthcare Technology Foundation (2006), *Impact of clinical alarms on patient safety.*

나선경, 이영희(2013), 중환자실 간호사의 패혈증에 대한 지식 조사, 『*Journal of Korean Clinical Nursing Research*』 V01.19(2013):455

이주리, 최혜란(2014), 일반병동 내 중증 패혈증 또는 패혈성 쇼크 환자의 중환자실 전동 예측에 대한 수정조기경고점수의 타당성, *대한간호학회지* 제44권 제2호(2014):220

정유진, 김현정(2018), 중환자실에서 사용되는 의료장비의 경보음 발생과 관리 현황, *Journal of Korean Biological Nursing Science* 2018;20(4):228-235

Barbara J. Drew and Patricia Harris(2014), *Insights into the problem of alarm fatigue with*

physiologic monitor devices: A comprehensive observational study of consecutive intensive care unit patients, PLOS ONE

Bojana Velichkovska, Hristijan Gjoreski, Daniel Denkovski, Marija Kalendar, Leo Anthnoy Celi and Venet Osmani (2021), *Investigating presence of ethnoracial bias in clinical data using machine learning*, Medrxiv.org

Borowski M, Görges M, Fried R, Such O, Wrede C and Imhoff M (2011). *Medical device alarms*, Biomedical Engineering/ Biomedizinische Technik. 2011;56(2):73-83.

Breiman and Leo (2001), *Random forests. machine learning*, 2001, 45.1: 5-32.

Chen, Tianqi, GUESTRIN and Carlos (2016), *XGBoost: A scalable tree boosting system.*, Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016. p. 785-794.

Chrystinne Fernandes, Simon Miles and Carlos José Pereira Lucena (2020), *Detecting false alarms by analyzing alarm-context information: Algorithm development and validation*, JMIR Med Inform 2020, May

David M. Maslove, Joel A. Dubin, Arvind Shrivat and Joon Lee (2016), *Errors, omissions, and outliers in hourly vital signs measurements in intensive care*, Online clinical Medicine

- Graham K and Cvach M (2010), *Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms.*, American Journal of Critical Care. 2010;19(1):28-34.
- H.J. Kam and H.Y. Kim (2017), *Learning representations for the early detection of sepsis with deep neural networks*, Comput. Biol. Med. 89 (2017) 248–255.
- Mohaimenul Islam, Tahmina Nasrin, Bruno Andreas Walther, Chieh-Chen Wu, Hsuan-Chia Yang and Yu-Chuan Li (2018), *Prediction of sepsis patients using machine learning approach: A meta-analysis*, Computer Methods and Programs in Biomedicine
- Ok Min Cho (2016), *Clinical alarms in intensive care units: Perceived obstacles of alarm management and alarm fatigue in nurses*, Healthcare Informatics Research, 2016 22(1):46-53
- Park MY (2016), *Alarm perception of nurse, alarm hazards on clinical alarms in intensive care units*, Daejeon: Chungnam National University; 2016.
- Q. Mao, M. Jay, J.L. Hoffman, J. Calvert, C. Barton, D. Shimabukuro, L. Shieh, U. Chettipally, G. Fletcher and Y. Kerem (2018), *Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU*, BMJ open 8 (2018) e017833
- Schmidhuber and Jürgen (2015), *Deep learning in neural networks: An overview. Neural networks*, 2015, 61: 85-117.

T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, Feldman and C. Barton (2016), *Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach*, JMIR Med. Inform. (2016) 4.

Tom J. Pollard, Alistair E. W. Johnson, Jesse D. Raffa, Leo A. Celi, Roger G. Mark and Omar Badawi (2018), *Data Descriptor: The eICU Collaborative Research Database, a freely available multi-center database for critical care research*, nature scientific data

Zheng Dai (2020), *Analysis of adult disease characteristics and mortality on MIMIC-III*, PLOS ONE

Eric K. Wei, Theodore Long and Mitchell H. Katz (2021), *Nine Lessons Learned From the COVID-19 Pandemic for Improving Hospital Care and Health Care Delivery*, <https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2782429>, JAMA Intern Med, July 23, 2021

Brad Brooks(2021), *TigerConnect and Critical Alert Join to Deliver Full-Spectrum Care Collaboration in the Cloud*, <https://tigerconnect.com/blog/tigerconnect-critical-alert-join-to-deliver-care-collaboration/>, 2022년 7월 18일,

바소프레신, <https://ko.wikipedia.org/wiki/바소프레신>, 2022년 7월 2일

청년 의사, 2021년 8월 19일, “코로나19 환자 급증... 생활치료센터도 의료진 부족
에 ‘허덕’”

부록

Septic Shock의 정의

Septic shock은 근본적인 순환 및 세포대사 이상이 사망률을 상당히 증가시킬 만큼 심각한 패혈증의 하위집합으로 정의된다. Sepsis 태스크포스는 septic shock를 심혈관 기능 장애와 구별하고 세포 대사 이상의 중요성을 인식하기 위한 광범위한 관점을 선호하였으며, 패혈증만 있는 것보다 패혈증 쇼크가 사망 가능성이 훨씬 높은 심각한 질환으로 반영되어야 한다는데 만장일치가 있었다.

Septic shock은 적절한 수액 치료에도 불구하고 mean arterial pressure \geq 65mmHg를 유지하기 위해 혈관수축제(vasopressor)가 필요하고 serum lactate level $>$ 2mmol/L (18mg/dL)인 지속적인 저혈압이 있는 패혈증의 임상 구성으로 정의된다. 이 기준으로는 병원 사망률이 40%가 넘어간다.

ICD-9 Code

국제 질병 분류(ICD)는 사망률 통계의 수집, 처리, 분류 및 제시에서 국제적 비교 가능성을 증진시키기 위해 고안된 임의의 질병에 대응되는 숫자 기반 코드이다. 각각의 질병마다 그에 대응하는 코드로 분류 및 할당되며, 이는 세계보건기구(WHO)가 발표한 ICD의 분류 구조, 선택 및 수정 규칙을 따른다.

ICD-9과 연관된 분류로 ICD-9-CM (Clinical Modification)이 존재한다.

ICD-9-CM은 ICD-9을 기반으로 하며 미국 national center for health statistics(NCHS)에서 만든 변형으로 입원, 외래 및 의사 사무실 활용과 관련된 진단 코드를 할당하는 데 사용된다. 국제적으로 사용되는 통계 목적의 질병 관련 코드이기 때문에 ICD-9 에는 패혈증, 중증 패혈증, Septic Shock에 대응되는 코드들이 존재하며, 이들의 전제 요소인 감염 관련 질병 코드들도 확인 가능하다.

Angus ICD-9-CM

2001년, Angus 박사 연구진은 임상 수정된 ICD-9-CM 코드를 사용하여 미국에서의 중증 패혈증 발병률 추정치를 발표하였다. 이 추정치는 감염을 식별하기 위한 ICD-9-CM 코드들과 급성 장기 기능장애를 식별하기 위한 코드로 구성된 ICD-9-CM 패혈증 기준을 통해 평가되었다.

```

/* Detect Infection Codes */
if substr(dx{i}, 1, 3) in
('001','002','003','004','005','008','009','010','011','012',
'013','014','015','016','017','018','020','021','022','023',
'024','025','026','027','030','031','032','033','034','035',
'036','037','038','039','040','041','090','091','092','093',
'094','095','096','097','098','100','101','102','103','104',
'110','111','112','114','115','116','117','118','320','322',
'324','325','420','421','451','461','462','463','464','465',
'481','482','485','486','494','510','513','540','541','542',
'566','567','590','597','601','614','615','616','681','682',
'683','686','730') or
substr(dx{i}, 1, 4) in
('5695','5720','5721','5750','5990','7110','7907',
'9966','9985','9993') or
substr(dx{i}, 1, 5) in
('49121','56201','56203','56211','56213','56983') then
infection = 1;
/* Detect Acute Organ Dysfunction Diagnosis Codes */
if substr(dx{i}, 1, 3) in
('458','293','570','584') or
substr(dx{i}, 1, 4) in
('7855','3483','3481','2874','2875','2869','2866','5734') then
organ_dysfunction = 1;
/* Detect explicit diagnosis of severe sepsis or septic shock */
if substr(dx{i}, 1, 5) in ('99592','78552') then
explicit_dx = 1;
end;

```

이러한 기준을 통해 감염과 급성 장기 기능장애를 모두 가진 환자를 확보하게 되고, 중증 패혈증 환자의 식별이 이루어진다. 그래서 연구자들은 역학 감시와 코호트 선택을 위한 중증 패혈증 환자의 대규모 코호트를 식별할 수 있었으며, 패혈증에 대해 가장 일반적으로 인용되는 통계로 자리잡았다.

차트 노트

차트 노트는 환자의 질병 상태, 최종 진단, 치료 과정 등을 적은 의무 기록지이다. 패혈증, septic shock이 적힌 차트 노트를 통해 패혈증 환자를 추출하였다.

Abstract

A study on reducing false-positive alarms in the intensive care unit using a machine learning model

Hyun, Jung Kyun

Seoul School of Integrated Sciences and Technologies

Advisor: Chang Joong Ho

There has been an incessant controversy over the necessity and subject of Internet contents regulation. Voluntary regulation by service providers' association is preferred to government regulation from the perspective of the nature of Internet, that is, free exchange of information.

This study proposed a meaningful alarm indicator method that can reduce the fatigue of medical staff due to the high false-positive alarm of the patient monitoring device in the medical environment in Korea, where there is a shortage of medical staff, especially in the intensive care unit.

Currently, the intensive care unit patient monitoring system is supposed to sound an alarm when the upper and lower limit values set for major biosignals are exceeded. As a method to help determine the authenticity of these false-positive alarms, a comparison with the patient's emergency state (in this study whether septic shock occurred) was considered

3 hours after the biosignal measurement time.

The data set used in the study was used by merging MIMIC-3 data and eICU data, which were collected from ICU data in the United States. The artificial intelligence algorithm used for shock prediction developed a predictive AI model using XGBoost, Random Forest, and Multi-layer Perceptron.

The method of verifying false-positive alarms for alarms is a true positive alarm or alarm condition if the patient's shock is recorded 3 hours later when the bio-signal at the time of measurement exceeds the threshold set for the bio-signal and the alarm is triggered. When the patient's shock occurrence was not recorded after 3 hours, it was defined as a false-positive alarm and the number of times was compared.

A total of 5,698 data were compared. Of these, 1,799 cases were cases where an alarm that exceeded the set threshold was sounded. In this case, septic shock was not recorded after 3 hours in 624 cases. This means a reduction of about 34.6%.

In the environment of the intensive care unit, there may be differences depending on the setting range of the warning range of the patient monitoring device or the operational method. It can be said that we saw the possibility that it would be possible as an auxiliary indicator.

The dataset used in this study is white-oriented data, and it is a dataset that may be different from the operation situation of the intensive care unit in Korea. If it can be seen, it is believed that it can be of some help to the domestic medical environment, which lacks medical staff.

Key words: prediction, false alarm, sepsis, ICU, machine learning

Student Number: 1854061516