



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0007793
(43) 공개일자 2023년01월13일

(51) 국제특허분류(Int. Cl.)

G16H 50/80 (2018.01) G06N 20/00 (2019.01)
G06Q 50/22 (2018.01) G16H 10/60 (2018.01)
G16H 50/20 (2018.01) G16H 50/30 (2018.01)
G16H 50/50 (2018.01)

(52) CPC특허분류

G16H 50/80 (2018.01)
G06N 20/00 (2021.08)

(21) 출원번호 10-2021-0088550

(22) 출원일자 2021년07월06일

심사청구일자 2021년07월06일

(71) 출원인

연세대학교 산학협력단

서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)

한국과학기술원

대전광역시 유성구 대학로 291(구성동)

(72) 발명자

구교철

서울특별시 용산구 서빙고로 35, 102동 1405호 (한강로3가, 용산시티파크1단지)

이광석

서울특별시 영등포구 국제금융로 108-6, B동 803호(여의도동, 진주아파트)

(뒷면에 계속)

(74) 대리인

이강욱, 김성훈

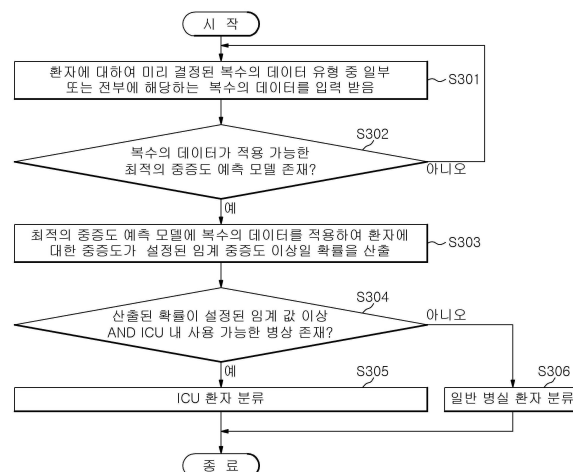
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 제한된 임상 데이터 및 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 최적의 중증도 예측 모델을 제공하기 위한 방법 및 장치

(57) 요약

본 발명은 제한된 임상 데이터 및 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 방법 및 장치에 관한 것이다. 구체적으로, 본 발명은 제한된 의료 자원 하에서 감염병 환자의 사망률을 낮추도록 최적 분류하기 위한 최적의 임계 값을 제공하고, 제한된 임상 데이터 하에서 감염병 환자를 최적 분류하기 위한 최적의 중증도 예측 모델을 제공하기 위한 방법 및 장치에 관한 것이다.

대표도 - 도3



(52) CPC특허분류

G06Q 50/22 (2021.08)
G16H 10/60 (2021.08)
G16H 50/20 (2018.01)
G16H 50/30 (2018.01)
G16H 50/50 (2018.01)

김정민

서울특별시 용산구 서빙고로 67, 102동 2406호 (용산동5가, 파크타워)

(72) 발명자

안재현

서울특별시 용산구 이촌로 347, 15동 905호 (서빙고동, 신동아아파트)

임하경

서울특별시 노원구 동일로208길 20, 208동 415호

이 발명을 지원한 국가연구개발사업

| | |
|-------------|---|
| 과제고유번호 | 1465033326 |
| 과제번호 | HW20C2103010021 |
| 부처명 | 보건복지부 |
| 과제관리(전문)기관명 | 한국보건산업진흥원 |
| 연구사업명 | 감염병방역기술개발(R&D) |
| 연구과제명 | COVID-19 의증 환자의 조기 선별진단을 위한 스마트 알고리즘 개발 |
| 기 여 율 | 1/1 |
| 과제수행기관명 | 연세대학교 산학협력단 |
| 연구기간 | 2021.03.01 ~ 2021.12.31 |

명세서

청구범위

청구항 1

전자 장치의 동작 방법에 있어서,

상기 전자 장치는 입력 장치, 출력 장치, 복수의 중증도 예측 모델이 저장된 메모리, 적어도 하나의 프로세서를 포함하고,

환자에 대하여 미리 결정된 복수의 데이터 유형 중 일부 또는 전부에 해당하는 복수의 데이터를 상기 입력 장치에 의하여 입력 받는 과정과,

상기 복수의 중증도 예측 모델 중 상기 복수의 데이터가 적용 가능한 최적의 모델을 상기 적어도 하나의 프로세서에 의하여 결정하는 과정과,

상기 최적의 모델에 상기 복수의 데이터를 적용하여 상기 환자에 대한 중증도가 설정된 임계 중증도 이상일 확률을 상기 적어도 하나의 프로세서에 의하여 산출하는 과정과,

상기 확률이 설정된 임계 값 이상인지 여부에 기반하여 상기 환자에 대한 환자 분류를 상기 적어도 하나의 프로세서에 의하여 결정하는 과정과,

상기 환자에 대한 상기 환자 분류를 상기 출력 장치에 의하여 출력하는 과정을 포함하는, 방법.

청구항 2

제1 항에 있어서,

상기 복수의 중증도 예측 모델은 상기 복수의 데이터 유형 중 일부 또는 전부에 기반하여 중증도가 설정된 임계 중증도 이상일 확률을 산출하도록 구성되고,

상기 복수의 중증도 예측 모델에 대한 우선 순위는 미리 설정되며,

상기 최적의 모델은 상기 복수의 데이터가 적용 가능한 적어도 하나의 중증도 예측 모델 중 설정된 우선 순위가 가장 높은 중증도 예측 모델인,

방법.

청구항 3

제2 항에 있어서,

상기 복수의 데이터 유형은 인구학적 데이터 (demographic data), 의학적 동반 질환 (medical comorbidities), 임상적 데이터 (clinical data) 및 실험실 데이터 (laboratory data)의 4가지 카테고리로 구성되고,

상기 인구학적 데이터 (demographic data)는 환자 나이 (patient age), 성별 (sex), 수축기 혈압 (systolic blood pressures), 이완기 혈압 (diastolic blood pressures), 심박수 (heart rate), 체온 (body temperature), 체질량 지수 (body mass index)의 7가지 데이터 유형을 포함하고,

상기 의학적 동반 질환 (medical comorbidities)은 고혈압 (hypertension), 당뇨병 (diabetes mellitus), 심부전 (heart failure), 심혈관 질환 (cardiovascular disease), 천식 (asthma), 만성 신장 질환 (chronic kidney disease), 만성 폐쇄성 폐 질환 (chronic obstructive pulmonary disease), 만성 간 질환 (chronic liver disease), 자가 면역 질환 (autoimmune disease), 치매 (dementia), 악성 종양 (malignancy), 임신 (pregnancy) 및 임신 주수 (pregnancy weeks)의 12가지 데이터 유형을 포함하고,

상기 임상적 데이터 (clinical data)는 발열 (fever), 기침 (cough), 객담 생성 (sputum production), 근육통 (myalgia), 피로 (fatigue), 인후통 (sore throat), 콧물 (rhinorrhea), 호흡 곤란 (shortness of breath), 구토 (vomiting), 설사 (diarrhea), 두통 (headache) 및 의식 변화 (altered consciousness)의 13가지 데이터 유형을 포함하고,

상기 실험실 데이터 (laboratory data)는 헤모글로빈 (hemoglobin), 헤마토크릿 (hematocrit), 백혈구 수 (white blood cell count), 림프구 (lymphocyte) 및 혈소판 수 (platelet count)의 5가지 데이터 유형을 포함하는,

방법.

청구항 4

제3 항에 있어서,

상기 복수의 증정도 예측 모델은 F+L+ 모델, F-L+ 모델, F+L- 모델, F-L- 모델로 구성되고,

상기 F+L+ 모델은 상기 복수의 데이터 유형에 포함되는 37가지 데이터 유형을 모두 사용하도록 구성되고,

상기 F-L+ 모델은 17가지 데이터 유형을 사용하도록 구성되고, 상기 17가지 데이터 유형은 상기 인구학적 데이터 (demographic data)에 포함되는 7가지 데이터 유형, 상기 실험실 데이터 (laboratory data)에 포함되는 5가지 데이터 유형, 호흡 곤란 (shortness of breath), 고혈압 (hypertension), 치매 (dementia), 당뇨병 (diabetes mellitus) 및 의식 변화 (altered consciousness)를 포함하고,

상기 F+L- 모델은 상기 37가지 데이터 유형 중 상기 실험실 데이터에 포함되는 5가지 데이터 유형을 제외한 32가지 데이터 유형을 사용하도록 구성되고,

상기 F-L- 모델은 11가지 데이터 유형을 사용하도록 구성되고, 상기 11가지 데이터 유형은 상기 인구학적 데이터 (demographic data)에 포함되는 7가지 데이터 유형, 호흡 곤란 (shortness of breath), 고혈압 (hypertension), 치매 (dementia) 및 당뇨병 (diabetes mellitus)을 포함하는,

방법.

청구항 5

제4 항에 있어서,

상기 복수의 증정도 예측 모델에 대한 우선 순위는,

F+L+ 모델, F-L+ 모델, F+L- 모델, F-L- 모델의 순서로 높은,

방법.

청구항 6

제1 항에 있어서,

중환자실에 사용 가능한 병상이 존재하는지 여부에 대한 정보를 상기 입력 장치에 의하여 입력 받는 과정을 더 포함하고,

상기 환자 분류는 중환자실 (intensive care unit, ICU) 환자 분류와 일반 병실 (general ward) 환자 분류로 구성되고,

상기 확률이 상기 설정된 임계 값 이상이고 중환자실에 사용 가능한 병상이 존재하는 경우, 상기 환자 분류는 중환자실 환자 분류로 결정되고,

상기 확률이 상기 설정된 임계 값 미만이거나 상기 중환자실에 사용 가능한 병상이 존재하지 않는 경우, 상기

환자 분류는 일반 병실 환자 분류로 결정되는,
방법.

청구항 7

전자 장치에 있어서,

상기 전자 장치는 입력 장치, 출력 장치, 복수의 중증도 예측 모델이 저장된 메모리, 적어도 하나의 프로세서를 포함하고,

제1 항 내지 제6 항 중 어느 한 항에 따른 방법을 수행하도록 구성된,

전자 장치.

청구항 8

제1 항 내지 제6 항 중 어느 한 항에 따른 방법을 수행하도록 구성되며, 컴퓨터 판독 가능한 저장 매체에 기록된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001]

본 발명은 제한된 임상 데이터 및 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 방법 및 장치에 관한 것이다. 구체적으로, 본 발명은 제한된 의료 자원 하에서 감염병 환자의 사망률을 낮추도록 최적 분류하기 위한 최적의 임계 값을 제공하고, 제한된 임상 데이터 하에서 감염병 환자를 최적 분류하기 위한 최적의 중증도 예측 모델을 제공하기 위한 방법 및 장치에 관한 것이다.

배경 기술

[0003]

2019 년 코로나 바이러스 질병 (coronavirus disease 2019, COVID-19)의 발병은 세계적인 유행병, 즉, 팬데믹 (pandemic)을 일으켰으며, 2021 년 4 월 현재 전세계적으로 150,110,310 명의 확진자 및 3,158,792 명의 사망자가 발생했다. COVID-19로 인한 높은 감염률, 중증 질병 발생 및 사망률은 세계 경제와 의료 시스템에 전례없는 부담이다. 이에 대한 대응으로 세계 보건기구 (World Health Organization, WHO)의 가이드 라인은 모든 국가가 호흡 보조 장치를 포함한 의료 시설의 급증에 대비하고 환자 분류 프로토콜을 구현할 것을 권장한다. 안타깝게도, WHO의 가이드 라인은 각 지역에 대한 발병 규모의 집계에 대하여 범용적인 한 가지 방법(one-size-fits-all approach)을 제공하지 못한다. 치료 장비의 획기적인 개선과 백신 접종을 통한 감염병에 대한 면역이 이루어지기 전까지, 다양한 환자 분류 임계 값 수준을 가진 적응형 환자 분류 모델은 제한된 의료 자원의 효과적인 활용을 위해 가장 중요할 것이다.

[0004]

확립된 자원이 부족하여 사회 경제적 혼란을 야기할 때, 급성 악화 및 사망 위험이 높은 임상적 스펙트럼을 가진 팬데믹 질병에 대한 다양한 전략적 환자 분류 조치가 필요하다. 첫째, 환자의 예후를 예측하기 위하여 신뢰할 수 있는 예측 모델은 선제적 치료와 후속 생존 결과 개선을 촉진할 수 있다. 이러한 예측 모델은 특정 임상 평가 변수에 초점을 맞출 뿐만 아니라 의사 결정에 도움이 될 수 있는 복합 질병 스펙트럼에 대한 예측도 제공해야 한다. 예를 들어, 기계적 환기가 필요한 질병 악화 위험이 높은 환자를 조기에 예측하면 임상가가 이러한 환자를 보다 적극적으로 모니터링하고 의료 자원의 사전 할당을 용이하게 할 수 있다. 둘째, 예측 모델은 쉽게 사용할 수 있는 평가 매개 변수에 따라 작동해야 하며, 자원이 제한된 환경에서와 같이 제한된 변수를 사용하여 신뢰성을 유지해야 한다. 더욱이, 특정 임상 평가 변수의 기여 요인에 대한 환자 별 해석 가능성을 제공하는 모델은 정확한 치료 계획에 도움이 될 수 있다. 셋째, 실제 모델은 동시 유입 환자의 양에 비해 시설 또는 국가 수준의 자원 가용성을 고려해야 한다. COVID-19는 대부분의 국가에서 의료 인프라의 붕괴와 관련이 있다. 따라서, 자원 가용성 궤도를 고려하고 실제로 입원과 집중 치료가 필요한 환자를 구별하는 조정 가능한 위험 계층화 모델은 의료 시스템의 부담을 줄일 수 있다.

[0005] COVID-19 환자를 효과적으로 분류하기 위해 수많은 예측 모델들이 개발되었다. 이러한 모델들은 보통의 예측 정확도를 보고했다. 그러나, 이러한 모델들의 신뢰성과 일반화 가능성은 단일 기관 데이터에 대한 의존, 단일 임상 결과 측정에 대한 제한, 데이터 부족시 차별적인 성능의 감소로 인해 의문을 제기했다. 가장 중요한 점으로서, 효과적인 자원 활용에 중요한 질병 중증도 예측 모델에 대한 환자 분류 임계 값은 제안된 모델들의 실제 구현 가능성을 제한할 수 있음에도 불구하고 무시된 주제였다. 자원 제한을 고려한 기존 시뮬레이션 연구는 COVID-19의 맥락에서 기계 학습 모델 또는 환자 유입 시나리오를 함께 활용하지 않았다. 종합하면, 이러한 모델들의 단점들은 다수의 대규모 기관 데이터를 기계 학습 및 시뮬레이션 모델링과 같은 고급 예측 모델과 결합해야 할 필요성을 강조한다.

[0006] 따라서, 본 발명은 제한된 의료 자원 하에서 감염병 환자의 사망률을 낮추도록 최적 분류하기 위한 최적의 임계 값을 제공하고, 제한된 임상 데이터 하에서 감염병 환자를 최적 분류하기 위한 최적의 중증도 예측 모델을 제공하기 위한 방법 및 장치가 필요한 실정이다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 등록특허 제10-1960504호(전염병 확산 예측 모델링 장치 및 방법)

발명의 내용

해결하려는 과제

[0009] 상술한 바와 같은 논의를 바탕으로, 본 발명의 다양한 실시 예들은 제한된 임상 데이터 및 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 방법 및 장치를 제공한다.

[0010] 또한, 본 발명의 다양한 실시 예들은 COVID-19 환자를 분류하기 위한 설명 특성이 있는 기준 예측 모델을 제공한다.

[0011] 또한, 본 발명의 다양한 실시 예들은 기준 예측 모델을 기반으로 재귀적 특성 제거 (recursive feature elimination, RFE) 기술을 사용하여 제한된 데이터 가용성 하에서 효율적인 활용을 위해 변수 제거 모델을 제공한다.

[0012] 또한, 본 발명의 다양한 실시 예들은 제한된 의료 자원 하에서 효율적인 자원 활용을 위해 이산 이벤트 시뮬레이션 (discrete event simulation, DES) 기술을 사용하여 적응형 분류 모델을 제공한다.

[0013] 또한, 본 발명의 다양한 실시 예들은 제한된 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 최적의 임계 값을 제공하기 위한 방법 및 장치를 제공한다.

[0014] 또한, 본 발명의 다양한 실시 예들은 제한된 임상 데이터 하에서 감염병 환자를 최적 분류하기 위한 최적의 중증도 예측 모델을 제공하기 위한 방법 및 장치를 제공한다.

과제의 해결 수단

[0016] 본 발명의 다양한 실시 예들에 따르면, 전자 장치의 동작 방법에 있어서, 전자 장치는 입력 장치, 출력 장치, 복수의 중증도 예측 모델이 저장된 메모리, 적어도 하나의 프로세서를 포함하고, 환자에 대하여 미리 결정된 복수의 데이터 유형 중 일부 또는 전부에 해당하는 복수의 데이터를 입력 장치에 의하여 입력 받는 과정과, 복수의 중증도 예측 모델 중 복수의 데이터가 적용 가능한 최적의 모델을 적어도 하나의 프로세서에 의하여 결정하는 과정과, 최적의 모델에 복수의 데이터를 적용하여 환자에 대한 중증도가 설정된 임계 중증도 이상일 확률을 적어도 하나의 프로세서에 의하여 산출하는 과정과, 확률이 설정된 임계 값 이상인지 여부에 기반하여 환자에 대한 환자 분류를 적어도 하나의 프로세서에 의하여 결정하는 과정과, 환자에 대한 환자 분류를 출력 장치에 의하여 출력하는 과정을 포함하는 방법이 제공된다.

[0017] 본 발명의 다양한 실시 예들에 따르면, 전자 장치에 있어서, 전자 장치는 입력 장치, 출력 장치, 복수의 중증도 예측 모델이 저장된 메모리, 적어도 하나의 프로세서를 포함하고, 본 발명의 다양한 실시 예들에 따른 전자 장치의 동작 방법을 수행하도록 구성된 전자 장치가 제공된다.

[0018] 본 발명의 다양한 실시 예들에 따르면, 컴퓨터 프로그램에 있어서, 본 발명의 다양한 실시 예들에 따른 전자 장치의 동작 방법을 수행하도록 구성되며, 컴퓨터 판독 가능한 저장 매체에 기록된 컴퓨터 프로그램이 제공된다.

발명의 효과

[0020] 본 발명의 다양한 실시 예들은 제한된 임상 데이터 및 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 방법 및 장치를 제공할 수 있다.

[0021] 또한, 본 발명의 다양한 실시 예들은 COVID-19 환자를 분류하기 위한 설명 특성이 있는 기준 예측 모델을 제공할 수 있다.

[0022] 또한, 본 발명의 다양한 실시 예들은 기준 예측 모델을 기반으로 재귀적 특성 제거 (recursive feature elimination, RFE) 기술을 사용하여 제한된 데이터 가용성 하에서 효율적인 활용을 위해 변수 제거 모델을 제공할 수 있다.

[0023] 또한, 본 발명의 다양한 실시 예들은 제한된 의료 자원 하에서 효율적인 자원 활용을 위해 이산 이벤트 시뮬레이션 (discrete event simulation, DES) 기술을 사용하여 적응형 분류 모델을 제공할 수 있다.

[0024] 또한, 본 발명의 다양한 실시 예들은 제한된 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 최적의 임계값을 제공하기 위한 방법 및 장치를 제공할 수 있다.

[0025] 또한, 본 발명의 다양한 실시 예들은 제한된 임상 데이터 하에서 감염병 환자를 최적 분류하기 위한 최적의 중증도 예측 모델을 제공하기 위한 방법 및 장치를 제공할 수 있다.

[0026] 본 발명에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0028] 도 1은 본 발명의 다양한 실시 예에 따라서 제한된 임상 데이터 및 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 전자 장치의 블록도를 도시한다.

도 2는 본 발명의 다양한 실시 예에 따라서 제한된 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 방법의 흐름도를 도시한다.

도 3은 본 발명의 다양한 실시 예에 따라서 제한된 임상 데이터 하에서 감염병 환자를 최적 분류하기 위한 방법의 흐름도를 도시한다.

도 4는 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델을 생성하기 위한 XGBoost에 기반한 의사결정 트리 (decision tree)의 일 예를 도시한다.

도 5는 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델을 생성하기 위한 훈련 코호트 및 홀드-아웃 코호트의 설명적 특성을 도시한다.

도 6은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델을 생성하기 위한 모델 별 성능 측정 결과를 도시한다.

도 7은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델에 대한 각 예측 변수의 샐플리 값 (shapley value)을 도시한다.

도 8은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델에 대한 각 예측 변수 별 샐플리 값 (shapley value)을 도시한다.

- 도 9는 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델의 상위 10개 예측 변수에 대한 실제 중증 환자 특정 SHAP 플롯을 도시한다.
- 도 10은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델의 상위 10개 예측 변수에 대한 실제 비-중증 환자 특정 SHAP 플롯을 도시한다.
- 도 11은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델 중 실험실 변수를 포함한 모델 (F+L+ model)의 AUROC (area under the receiver operating characteristic)를 도시한다.
- 도 12는 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델 중 실험실 변수를 제외한 모델 (F+L-)의 AUROC (area under the receiver operating characteristic)를 도시한다.
- 도 13은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델 중 실험실 변수를 포함한 모델 (F+L+ model) 및 실험실 변수를 제외한 모델 (F+L-)에 대하여 예측 변수의 개수에 따른 성능을 도시한다.
- 도 14는 본 발명의 다양한 실시 예들에 따른 각각의 중증도 예측 모델 별 성능을 도시한다.
- 도 15는 본 발명의 다양한 실시 예들에 따른 각각의 중증도 예측 모델 별 과적합(over-fitting)을 방지하기 위한 검증 결과를 도시한다.
- 도 16는 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델에 대하여 재귀적 특성 감소 (recursive feature elimination, RFE)를 적용하기 위한 예측 변수의 순위를 도시한다.
- 도 17은 본 발명의 다양한 실시 예들에 따른 제한된 의료 자원 가용성 하에서 환자 사망률 시뮬레이션을 생성하기 위한 워크 플로우를 도시한다.
- 도 18은 본 발명의 다양한 실시 예들에 따른 환자 유입 시뮬레이션을 생성하기 위한 과거 환자 유입 시나리오를 도시한다.
- 도 19는 본 발명의 다양한 실시 예들에 따른 기초 감염재생산 지수 (basic reproduction rate) (R_0)에 기반한 가상의 환자 유입 시나리오를 도시한다.
- 도 20은 본 발명의 다양한 실시 예들에 따른 중증 환자 유입 시뮬레이션을 생성하기 위한 중증 및 중증이 아닌 환자의 예측 확률 분포와 생성된 예측 확률 분포를 도시한다.
- 도 21은 본 발명의 다양한 실시 예들에 따른 과거 역사적 환자 유입 시나리오에서 중증도 판단의 임계 확률에 따른 사망률의 그래프를 도시한다.
- 도 22는 본 발명의 다양한 실시 예들에 따른 과거 역사적 환자 유입 시나리오에서 중증도 판단의 임계 확률에 따른 사망 하위 유형 별로 분해된 사망률의 그래프를 도시한다.
- 도 23은 본 발명의 다양한 실시 예들에 따른 과거 역사적 환자 유입 시나리오 별 최적 임계 확률, 최적화된 사망률, 감소된 사망률을 도시한다.
- 도 24는 본 발명의 다양한 실시 예들에 따른 기초 감염재생산 지수 (basic reproduction rate) (R_0)에 기반한 가상의 환자 유입 시나리오에서 중증도 판단의 임계 확률에 따른 사망률의 그래프를 도시한다.
- 도 25는 본 발명의 다양한 실시 예들에 따른 기초 감염재생산 지수 (basic reproduction rate) (R_0)에 기반한 가상의 환자 유입 시나리오에서 중증도 판단의 임계 확률에 따른 사망 하위 유형 별로 분해된 사망률의 그래프를 도시한다.
- 도 26은 본 발명의 다양한 실시 예들에 따른 기초 감염재생산 지수 (basic reproduction rate) (R_0)에 기반한 가상의 환자 유입 시나리오 별 최적 임계 확률, 최적화된 사망률, 감소된 사망률을 도시한다.
- 도 27은 본 발명의 다양한 실시 예들에 따른 각각의 과거 역사적 환자 유입 시나리오 및 기초 감염재생산 지수 (basic reproduction rate) (R_0)에 기반한 가상의 환자 유입 시나리오 별 최적 임계 확률, 최적화된 사망률, 감소된 사망률을 도시한다.

발명을 실시하기 위한 구체적인 내용

[0029]

본 발명에서 사용되는 용어들은 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 다른 실시 예의 범위를 한정하려는 의도가 아닐 수 있다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함

할 수 있다. 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 용어들은 본 발명에 기재된 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가질 수 있다. 본 발명에 사용된 용어들 중 일반적인 사전에 정의된 용어들은, 관련 기술의 문맥상 가지는 의미와 동일 또는 유사한 의미로 해석될 수 있으며, 본 발명에서 명백하게 정의되지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다. 경우에 따라서, 본 발명에서 정의된 용어일지라도 본 발명의 실시 예들을 배제하도록 해석될 수 없다.

- [0031] 도 1은 본 발명의 다양한 실시 예에 따라서 제한된 임상 데이터 및 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 전자 장치의 블록도를 도시한다.
- [0032] 도 1을 참조하면, 본 발명의 다양한 실시 예에 따라서 제한된 임상 데이터 및 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 전자 장치(100)는 입력 장치(110), 출력 장치(120), 메모리(130), 프로세서(140)를 포함한다.
- [0033] 입력 장치(110)는 프로세서(140)와 연결되고 정보 등을 입력할 수 있다. 일 실시 예에 따라서, 입력 장치(110)는 도시되지 않았으나 전자 장치(100)에 추가로 포함될 수 있는 송수신기를 통해 통신 네트워크로 연결된 다른 장치로부터 수신한 정보 등을 입력할 수 있다. 입력 장치(110)는 전자 장치(100)의 터치 디스플레이, 키 패드, 키보드 등을 포함할 수 있다.
- [0034] 출력 장치(120)는 프로세서(140)와 연결되고 정보 등을 영상/음성 등의 형태로 출력할 수 있다. 일 실시 예에 따라서, 출력 장치(140)는 입력 장치(110)를 통해 입력된 정보, 프로세서(140)에 의하여 생성된 정보, 도시되지 않았으나 전자 장치(100)에 추가로 포함될 수 있는 송수신기를 통해 통신 네트워크로 연결된 다른 장치로부터 수신한 정보 등을 출력할 수 있다. 출력 장치(140)는 전자 장치(100)의 디스플레이, 스피커 등을 포함할 수 있다.
- [0035] 메모리(130)는 프로세서(140)와 연결되고, 프로세서(140)의 동작을 위한 기본 프로그램, 응용 프로그램, 설정 정보, 입력 장치(110)를 통해 입력된 정보, 프로세서(140)의 연산에 의하여 생성된 정보 등의 데이터를 저장할 수 있다. 또한, 메모리(130)는 도시되지 않았으나 전자 장치(100)에 추가로 포함될 수 있는 송수신기와 연결되고 통신을 통해 수신한 이미지 및 정보 등을 저장할 수 있다. 메모리(130)는 휘발성 메모리, 비휘발성 메모리 또는 휘발성 메모리와 비휘발성 메모리의 조합으로 구성될 수 있다. 그리고, 메모리(130)는 프로세서(140)의 요청에 따라 저장된 데이터를 제공할 수 있다.
- [0036] 프로세서(140)는 본 발명에서 제안한 절차 및/또는 방법들을 구현하도록 구성될 수 있다. 프로세서(140)는 전자 장치(100)의 전반적인 동작들을 제어한다. 예를 들어, 프로세서(140)는 입력 장치(110)를 통해 정보를 입력 받는다. 또한, 프로세서(140)는 출력 장치(120)를 통해 정보를 출력한다. 또한, 프로세서(140)는 메모리(130)에 데이터를 기록하고, 읽는다. 또한, 프로세서(140)는 도시되지 않았으나 전자 장치(100)에 추가로 포함될 수 있는 송수신기를 통해 정보 등을 다른 장치와 전송 또는 수신한다. 프로세서(140)는 적어도 하나의 프로세서(processor)를 포함할 수 있다.
- [0038] 도 2는 본 발명의 다양한 실시 예에 따라서 제한된 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 방법의 흐름도를 도시한다. 도 2의 실시 예에서, 전자 장치는 입력 장치, 출력 장치, 중증도 예측 모델이 저장된 메모리, 적어도 하나의 프로세서를 포함한다.
- [0039] 도 2를 참조하면, S201 단계에서, 전자 장치는 기초 감염재생산 지수 (basic reproduction rate) (R0)를 포함하는 기초 설정을 상기 입력 장치에 의하여 입력 받는다.
- [0040] S202 단계에서, 전자 장치는 기초 설정을 SIR (감염가능자 (susceptible) - 감염자 (infected) - 회복자 (recovered) model) 모델에 적용하여 환자 유입 및 중증 환자 유입 시뮬레이션을 상기 적어도 하나의 프로세서에 의하여 생성한다.
- [0041] 일 실시 예에 따르면, SIR 모델은 다음의 [수학식 1]으로 구성되며,
- [0042] [수학식 1]
- [0043]
$$dS/dt = -\beta \cdot S \cdot I/N$$

- [0044] $dI/dt = \beta \cdot S \cdot I/N - \gamma \cdot I$
- [0045] $dR/dt = \gamma \cdot I$
- [0046] 상기 [수학식 1]에서, β 는 감염률이고, γ 는 회복률이고, S 는 감염 가능한 사람의 수이고, I 는 감염된 사람의 수이고, R 은 회복된 사람의 수이고, t 는 시간이고, $N=S(t)+I(t)+R(t)$ 이고 고정된 총 인구 수를 나타내며, β 는 상기 R_0 에 기반하여 소정의 범위에서 결정된 값으로 구성될 수 있다.
- [0047] 일 실시 예에 따르면, $N = S(t) + I(t) + R(t) = 60,000$, $I(0) = 6$, $R(0) = 0$, $\gamma = 0.05$ 으로 설정되고, β 는 상기 R_0 에 기반하여 0.75 이상 5 이하의 범위에서 결정된 값으로 설정될 수 있다.
- [0048] S203 단계에서, 전자 장치는 환자 유입 및 중증 환자 유입 시뮬레이션에 대하여 제한된 중환자실 (intensive care unit, ICU) 용량을 적용함으로써 임계 값에 따른 사망률 시뮬레이션을 상기 적어도 하나의 프로세서에 의하여 생성한다. 일 실시 예에 따라서, 사망률 시뮬레이션은, 특정 환자의 중증도가 상기 임계 중증도 이상일 확률이 상기 임계 값 이상이고 상기 제한된 ICU 용량이 수용 가능한 경우 ICU에 입원하고, 상기 특정 환자의 중증도가 상기 임계 중증도 이상일 확률이 상기 임계 값 미만이거나 또는 상기 제한된 ICU 용량이 수용 불가능한 경우 일반 병실 (general ward)로 입원하며, 상기 특정 환자의 실제 중증도가 상기 임계 중증도 이상이고 ICU에 입원할 경우 상기 특정 환자의 사망률이 50%이고, 상기 특정 환자의 실제 중증도가 상기 임계 중증도 이상이고 일반 병실에 입원할 경우 상기 특정 환자의 사망률이 99%라는 설정에 기반하여 생성된다.
- [0049] 일 실시 예에 따르면, 제한된 ICU 용량은 504로 설정될 수 있다.
- [0050] S204 단계에서, 전자 장치는 사망률 시뮬레이션에서 사망률이 최소화되는 최적 임계 값을 상기 적어도 하나의 프로세서에 의하여 결정한다. 일 실시 예에 따르면, 사망률은 (총 사망자 수) / (총 환자 수)이다.
- [0051] 일 실시 예에 따르면, 상기 최적 임계 값은 상기 사망률 시뮬레이션에서 상기 사망률이 최소화되는 임계 값이며, 또한, 상기 최적 임계 값은 상기 사망률 시뮬레이션에서 전체 사망 중 I 유형 사망의 비율이 최대로 되는 임계 값이고, 상기 전체 사망은 I 유형 사망, II 유형 사망, III 유형 사망으로 구성되며, I 유형 사망은 중증 환자가 ICU로 이송하였으나 사망하는 유형이고, II 유형 사망은 중증 환자로 분류되었으나 ICU 용량이 제한되어 일반 병실로 이송되어 사망하는 유형이고, III 유형 사망은 중증 환자가 비-중증 환자로 잘못 분류되고 일반 병실로 이송되어 사망하는 유형이다.
- [0052] S205 단계에서, 전자 장치는 최적 임계 값을 상기 중증도 예측 모델에 적용하여 환자에 대한 환자 분류를 상기 적어도 하나의 프로세서에 의하여 수행한다. 또한, 전자 장치는 상기 환자에 대한 상기 환자 분류를 상기 출력 장치에 의하여 출력한다. 일 실시 예에 따르면, 환자 분류는 ICU 환자 분류와 일반 병실 환자 분류로 구성된다.
- [0053] 일 실시 예에 따르면, 최적 임계 값에 기반하여 환자 분류를 수행하는 과정은, 환자에 대하여 미리 결정된 복수의 데이터 유형 중 일부 또는 전부에 해당하는 복수의 데이터를 상기 입력 장치에 의하여 입력 받는 과정과, 상기 중증도 예측 모델에 상기 복수의 데이터를 적용하여 상기 환자에 대한 중증도가 상기 임계 중증도 이상일 중증도 확률을 상기 적어도 하나의 프로세서에 의하여 산출하는 과정과, 상기 중증도 확률이 상기 최적 임계 값 이상인지 여부에 기반하여 상기 환자에 대한 상기 환자 분류를 상기 적어도 하나의 프로세서에 의하여 결정하는 과정을 포함할 수 있다.
- [0054] 일 실시 예에 따르면, 상기 복수의 데이터 유형은 인구학적 데이터 (demographic data), 의학적 동반 질환 (medical comorbidities), 임상적 데이터 (clinical data) 및 실험실 데이터 (laboratory data)의 4가지 카테고리 구성되고,
- [0055] 상기 인구학적 데이터 (demographic data)는 환자 나이 (patient age), 성별 (sex), 수축기 혈압 (systolic blood pressures), 이완기 혈압 (diastolic blood pressures), 심박수 (heart rate), 체온 (body temperature), 체질량 지수 (body mass index)의 7가지 데이터 유형을 포함하고,
- [0056] 상기 의학적 동반 질환 (medical comorbidities)은 고혈압 (hypertension), 당뇨병 (diabetes mellitus), 심부전 (heart failure), 심혈관 질환 (cardiovascular disease), 천식 (asthma), 만성 신장 질환 (chronic kidney disease), 만성 폐쇄성 폐 질환 (chronic obstructive pulmonary disease), 만성 간 질환 (chronic liver disease), 자가 면역 질환 (autoimmune disease), 치매 (dementia), 악성 종양 (malignancy), 임신 (pregnancy) 및 임신 주수 (pregnancy weeks)의 12가지 데이터 유형을 포함하고,
- [0057] 상기 임상적 데이터 (clinical data)는 발열 (fever), 기침 (cough), 객담 생성 (sputum production), 근육통

(myalgia), 피로 (fatigue), 인후통 (sore throat), 콧물 (rhinorrhea), 호흡 곤란 (shortness of breath), 구토 (vomiting), 설사 (diarrhea), 두통 (headache) 및 의식 변화 (altered consciousness)의 13가지 데이터 유형을 포함하고,

- [0058] 상기 실험실 데이터 (laboratory data)는 헤모글로빈 (hemoglobin), 헤마토크릿 (hematocrit), 백혈구 수 (white blood cell count), 림프구 (lymphocyte) 및 혈소판 수 (platelet count)의 5가지 데이터 유형을 포함할 수 있다.
- [0059] 일 실시 예에 따르면, 도 2의 실시 예에 따른 방법은, ICU에 사용 가능한 병상이 존재하는지 여부에 대한 정보를 상기 입력 장치에 의하여 입력 받는 과정을 더 포함하고, 상기 중증도 확률이 상기 최적 임계 값 이상이고 중환자실에 사용 가능한 병상이 존재하는 경우, 상기 환자 분류는 상기 ICU 환자 분류로 결정되고, 상기 중증도 확률이 상기 최적 임계 값 미만이거나 상기 중환자실에 사용 가능한 병상이 존재하지 않는 경우, 상기 환자 분류는 상기 일반 병실 환자 분류로 결정될 수 있다.
- [0060] 일 실시 예에 따르면, 전자 장치에 있어서, 상기 전자 장치는 입력 장치, 출력 장치, 복수의 중증도 예측 모델이 저장된 메모리, 적어도 하나의 프로세서를 포함하고, 상술한 도 2의 실시 예에 따른 방법을 수행하도록 구성된 전자 장치가 제공될 수 있다.
- [0061] 일 실시 예에 따르면, 컴퓨터 프로그램에 있어서, 상술한 도 2의 실시 예에 따른 방법을 수행하도록 구성되며, 컴퓨터 판독 가능한 저장 매체에 기록된 컴퓨터 프로그램이 제공될 수 있다.
- [0063] 도 3은 본 발명의 다양한 실시 예에 따라서 제한된 임상 데이터 하에서 감염병 환자를 최적 분류하기 위한 방법의 흐름도를 도시한다. 도 3의 실시 예에서, 전자 장치는 입력 장치, 출력 장치, 복수의 중증도 예측 모델이 저장된 메모리, 적어도 하나의 프로세서를 포함한다.
- [0064] 도 3을 참조하면, S301 단계에서, 전자 장치는 환자에 대하여 미리 결정된 복수의 데이터 유형 중 일부 또는 전부에 해당하는 복수의 데이터를 상기 입력 장치에 의하여 입력 받는다.
- [0065] 일 실시 예에 따르면, 상기 복수의 데이터 유형은 인구학적 데이터 (demographic data), 의학적 동반 질환 (medical comorbidities), 임상적 데이터 (clinical data) 및 실험실 데이터 (laboratory data)의 4가지 카테고리 구성되고,
- [0066] 상기 인구학적 데이터 (demographic data)는 환자 나이 (patient age), 성별 (sex), 수축기 혈압 (systolic blood pressures), 이완기 혈압 (diastolic blood pressures), 심박수 (heart rate), 체온 (body temperature), 체질량 지수 (body mass index)의 7가지 데이터 유형을 포함하고,
- [0067] 상기 의학적 동반 질환 (medical comorbidities)은 고혈압 (hypertension), 당뇨병 (diabetes mellitus), 심부전 (heart failure), 심혈관 질환 (cardiovascular disease), 천식 (asthma), 만성 신장 질환 (chronic kidney disease), 만성 폐쇄성 폐 질환 (chronic obstructive pulmonary disease), 만성 간 질환 (chronic liver disease), 자가 면역 질환 (autoimmune disease), 치매 (dementia), 악성 종양 (malignancy), 임신 (pregnancy) 및 임신 주수 (pregnancy weeks)의 12가지 데이터 유형을 포함하고,
- [0068] 상기 임상적 데이터 (clinical data)는 발열 (fever), 기침 (cough), 객담 생성 (sputum production), 근육통 (myalgia), 피로 (fatigue), 인후통 (sore throat), 콧물 (rhinorrhea), 호흡 곤란 (shortness of breath), 구토 (vomiting), 설사 (diarrhea), 두통 (headache) 및 의식 변화 (altered consciousness)의 13가지 데이터 유형을 포함하고,
- [0069] 상기 실험실 데이터 (laboratory data)는 헤모글로빈 (hemoglobin), 헤마토크릿 (hematocrit), 백혈구 수 (white blood cell count), 림프구 (lymphocyte) 및 혈소판 수 (platelet count)의 5가지 데이터 유형을 포함할 수 있다.
- [0070] S302 단계에서, 전자 장치는 복수의 중증도 예측 모델 중 상기 복수의 데이터가 적용 가능한 최적의 모델을 상기 적어도 하나의 프로세서에 의하여 결정한다. 일 실시 예에 따르면, 전자 장치는 입력된 복수의 데이터가 적용 가능한 최적의 모델을 결정함으로써 최적의 모델이 존재하는 경우 S303 단계로 진행하지만, 최적의 모델을 결정하지 못하여 최적의 모델이 존재하지 않는 경우 S301 단계로 돌아갈 수 있다.
- [0071] 일 실시 예에 따르면, 상기 복수의 중증도 예측 모델은 상기 복수의 데이터 유형 중 일부 또는 전부에 기반하여

중증도가 설정된 임계 중증도 이상일 확률을 산출하도록 구성되고, 상기 복수의 중증도 예측 모델에 대한 우선 순위는 미리 설정되며, 상기 최적의 모델은 상기 복수의 데이터가 적용 가능한 적어도 하나의 중증도 예측 모델 중 설정된 우선 순위가 가장 높은 중증도 예측 모델일 수 있다. 일 실시 예에 따르면, 상기 복수의 중증도 예측 모델에 대한 우선 순위는 각 중증도 예측 모델의 AUROC가 높은 순위에 따라서 설정될 수 있다. 일 실시 예에 따르면, 상기 복수의 중증도 예측 모델에 대한 우선 순위는 F+L+ 모델, F-L+ 모델, F+L- 모델, F-L- 모델의 순서로 높을 수 있다.

- [0072] 일 실시 예에 따르면, 상기 복수의 중증도 예측 모델은 F+L+ 모델, F-L+ 모델, F+L- 모델, F-L- 모델로 구성되고,
- [0073] 상기 F+L+ 모델은 상기 복수의 데이터 유형에 포함되는 37가지 데이터 유형을 모두 사용하도록 구성되고,
- [0074] 상기 F-L+ 모델은 17가지 데이터 유형을 사용하도록 구성되고, 상기 17가지 데이터 유형은 상기 인구학적 데이터 (demographic data)에 포함되는 7가지 데이터 유형, 상기 실험실 데이터 (laboratory data)에 포함되는 5가지 데이터 유형, 호흡 곤란 (shortness of breath), 고혈압 (hypertension), 치매 (dementia), 당뇨병 (diabetes mellitus) 및 의식 변화 (altered consciousness)를 포함하고,
- [0075] 상기 F+L- 모델은 상기 37가지 데이터 유형 중 상기 실험실 데이터에 포함되는 5가지 데이터 유형을 제외한 32가지 데이터 유형을 사용하도록 구성되고,
- [0076] 상기 F-L- 모델은 11가지 데이터 유형을 사용하도록 구성되고, 상기 11가지 데이터 유형은 상기 인구학적 데이터 (demographic data)에 포함되는 7가지 데이터 유형, 호흡 곤란 (shortness of breath), 고혈압 (hypertension), 치매 (dementia) 및 당뇨병 (diabetes mellitus)을 포함할 수 있다.
- [0077] S303 단계에서, 전자 장치는 최적의 모델에 상기 복수의 데이터를 적용하여 상기 환자에 대한 중증도가 설정된 임계 중증도 이상일 확률을 상기 적어도 하나의 프로세서에 의하여 산출한다.
- [0078] S304 단계에서, 전자 장치는 확률이 설정된 임계 값 이상인지 여부에 기반하여 상기 환자에 대한 환자 분류를 상기 적어도 하나의 프로세서에 의하여 결정한다. 일 실시 예에 따르면, 전자 장치는 산출된 확률이 설정된 임계 값 이상인 조건이 만족됨과 동시에 ICU 내 사용 가능한 병상이 존재한다는 조건이 만족되는 경우 환자에 대하여 ICU 환자 분류를 결정하고 S305 단계로 진행하지만, 두 조건 중 어느 하나라도 만족되지 않는 경우 환자에 대하여 일반 병실 환자 분류를 결정하고 S306 단계로 진행할 수 있다.
- [0079] S305 단계에서, 전자 장치는 환자에 대한 ICU 환자 분류를 출력 장치에 의하여 출력한다.
- [0080] S306 단계에서, 전자 장치는 환자에 대한 일반 병실 환자 분류를 상기 출력 장치에 의하여 출력한다.
- [0081] 일 실시 예에 따르면, 도 3의 실시 예에 따른 방법은 중환자실에 사용 가능한 병상이 존재하는지 여부에 대한 정보를 상기 입력 장치에 의하여 입력 받는 과정을 더 포함하고, 상기 환자 분류는 중환자실 (intensive care unit, ICU) 환자 분류와 일반 병실 (general ward) 환자 분류로 구성되고, 상기 확률이 상기 설정된 임계 값 이상이고 중환자실에 사용 가능한 병상이 존재하는 경우, 상기 환자 분류는 중환자실 환자 분류로 결정되고, 상기 확률이 상기 설정된 임계 값 미만이거나 상기 중환자실에 사용 가능한 병상이 존재하지 않는 경우, 상기 환자 분류는 일반 병실 환자 분류로 결정될 수 있다.
- [0082] 일 실시 예에 따르면, 전자 장치에 있어서, 상기 전자 장치는 입력 장치, 출력 장치, 복수의 중증도 예측 모델이 저장된 메모리, 적어도 하나의 프로세서를 포함하고, 상술한 도 3의 실시 예에 따른 방법을 수행하도록 구성된 전자 장치가 제공될 수 있다.
- [0083] 일 실시 예에 따르면, 컴퓨터 프로그램에 있어서, 상술한 도 3의 실시 예에 따른 방법을 수행하도록 구성되며, 컴퓨터 판독 가능한 저장 매체에 기록된 컴퓨터 프로그램이 제공될 수 있다.
- [0085] 도 4는 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델을 생성하기 위한 XGBoost에 기반한 의사결정 트리 (decision tree)의 일 예를 도시한다.
- [0086] 의사결정 트리 (decision tree)는 분류 (classification)와 회귀 (regression)가 모두 가능한 지도 학습 모델 중 하나이다.
- [0087] 그래디언트 부스팅 알고리즘 (gradient boosting algorithm, GBM)은 의사결정 트리 (decision tree)를 기반으

로 한 기계 학습(machine learning) 모델이다. 그래디언트 부스팅 알고리즘은 회귀분석 또는 분류 분석을 수행할 수 있는 예측모형이며, 예측모형의 앙상블 방법론 중 부스팅 계열에 속하는 알고리즘이다.

[0088] XGBoost(extreme gradient boosting)는 속도, 세부적인 기능 차이 등에 있어서 그래디언트 부스팅 알고리즘의 변형 알고리즘이다. XGBoost는 의사결정 트리 부스팅을 위한 고효율의 확장 가능한 기계학습 기법으로, 기존 의사결정 트리 부스팅 기반의 기법보다 높은 성능을 보여 다양한 분야에서 사용되고 있다. XGBoost는 의사결정 트리 앙상블 기법 중 하나로, 의사결정 트리를 구성할 때 병렬 처리 기법을 사용하기 때문에 수행시간이 빠르다는 장점이 있다.

[0089] XGBoost는 병렬 처리로 학습, 분류 속도가 빠르기 때문에 GBM 대비 빠른 수행시간을 갖는다. 또한, 표준 GBM의 경우 과적합 규제기능이 없으나, XGBoost는 자체에 과적합 규제 기능으로 강한 내구성을 지닌다. 또한, XGBoost는 CART(Classification and regression tree) 앙상블 모델을 사용하여 분류와 회귀영역에서 뛰어난 예측 성능을 발휘한다.

[0090] 도 4는 XGBoost에 기반한 의사결정 트리의 일 예를 도시한다.

[0091] 본 발명의 다양한 실시 예들에 따른 전자 장치는 XGBoost에 기반하여 중증도 예측 모델을 생성할 수 있다. 또한, 본 발명의 다양한 실시 예들에 따른 전자 장치는 XGBoost에 기반하여 생성된 중증도 예측 모델을 이용하여 제한된 의료 자원 하에서 감염병 환자의 사망률을 낮추도록 최적 분류하기 위한 최적의 임계 값을 제공하고, 제한된 임상 데이터 하에서 감염병 환자를 최적 분류하기 위한 최적의 중증도 예측 모델을 제공할 수 있다.

[0093] 이하, 제한된 임상 데이터 및 의료 자원 하에서 감염병 환자를 최적 분류하기 위한 방법 및 장치, 구체적으로, 제한된 임상 데이터 및 의료 자원 하에서 COVID-19 환자의 최적 분류하기 위한 장치 및 방법을 개발하기 위한 연구를 설명한다.

[0095] 요약

[0096] 2019 년 코로나 바이러스 질병(coronavirus disease 2019, COVID-19)의 세계적인 유행은 의료 시스템에 전례 없는 부담을 주고 있다. 자원 가용성을 차지하는 악화 위험이 높은 환자의 분류는 자원 제약 환경에서 효율적인 임상 의사 결정 및 의료 시스템 관리에 최적이다. 5601 명의 환자의 임상 데이터를 사용하여 XGBoost를 사용하여 0.97의 AUROC를 달성한 예측 모델을 개발했다. 예측 모델의 리소스 제약 환경에서의 적용을 고려하여, 본 출원인은 예측 모델의 예측 성능을 유지하면서 반복적 특성 감소 기술로 예측 모델을 개선했다. 또한, 본 출원인은 분류 임계 값 수준을 조정하여 의료 시스템 용량이 유지되는 동안 환자 사망률을 최소화할 수 있도록 다양한 환자 유입 시나리오에서 사망률 제약을 예측하는 이산 이벤트 시뮬레이션 모델을 개발했다. 본 연구는 COVID-19의 향후 관리가 최대 환자 치료 효과를 보장하기 위해 의료 분야 및 헬스케어 관리 분야 간에 통합될 수 있음을 의미한다.

[0098] 소개 (Introduction)

[0099] 2019 년 코로나 바이러스 질병(coronavirus disease 2019, COVID-19)의 발병은 세계적인 유행병, 즉, 팬데믹(pandemic)을 일으켰으며, 2021 년 4 월 현재 전세계적으로 150,110,310 명의 확진자 및 3,158,792 명의 사망자가 발생했다. COVID-19로 인한 높은 감염률, 중증 질병 발생 및 사망률은 세계 경제와 의료 시스템에 전례 없는 부담이다. 이에 대한 대응으로 세계 보건기구(World Health Organization, WHO)의 가이드 라인은 모든 국가가 호흡 보조 장치를 포함한 의료 시설의 급증에 대비하고 환자 분류 프로토콜을 구현할 것을 권장한다. 안타깝게도, WHO의 가이드 라인은 각 지역에 대한 발병 규모의 집계에 대하여 범용적인 한 가지 방법(one-size-fits all approach)을 제공하지 못한다. 치료 장비의 획기적인 개선과 백신 접종을 통한 감염병에 대한 면역이 이루어지기 전까지, 다양한 환자 분류 임계 값 수준을 가진 적응형 환자 분류 모델은 제한된 의료 자원의 효과적인 활용을 위해 가장 중요할 것이다.

[0100] 확립된 자원이 부족하여 사회 경제적 혼란을 야기할 때, 급성 악화 및 사망 위험이 높은 임상적 스펙트럼을 가진 팬데믹 질병에 대한 다양한 전략적 환자 분류 조치가 필요하다. 첫째, 환자의 예후를 예측하기 위하여 신뢰할 수 있는 예측 모델은 선제적 치료와 후속 생존 결과 개선을 촉진할 수 있다. 이러한 예측 모델은 특정 임상

평가 변수에 초점을 맞출 뿐만 아니라 의사 결정에 도움이 될 수 있는 복합 질병 스펙트럼에 대한 예측도 제공해야 한다. 예를 들어, 기계적 환기가 필요한 질병 악화 위험이 높은 환자를 조기에 예측하면 임상의가 이러한 환자를 보다 적극적으로 모니터링하고 의료 자원의 사전 할당을 용이하게 할 수 있다. 둘째, 예측 모델은 쉽게 사용할 수 있는 평가 매개 변수에 따라 작동해야 하며, 자원이 제한된 환경에서와 같이 제한된 변수를 사용하여 신뢰성을 유지해야 한다. 더욱이, 특정 임상 평가 변수의 기여 요인에 대한 환자 특정 해석 가능성을 제공하는 모델은 정확한 치료 계획에 도움이 될 수 있다. 셋째, 실제 모델은 동시 유입 환자의 양에 비해 시설 또는 국가 수준의 자원 가용성을 고려해야 한다. COVID-19는 대부분의 국가에서 의료 인프라의 붕괴와 관련이 있다. 따라서, 자원 가용성 궤도를 고려하고 실제로 입원과 집중 치료가 필요한 환자를 구별하는 조정 가능한 위험 계층화 모델은 의료 시스템의 부담을 줄일 수 있다.

[0101] COVID-19 환자를 효과적으로 분류하기 위해 수많은 예측 모델들이 개발되었다. 이러한 모델들은 보통의 예측 정확도를 보고했다. 그러나, 이러한 모델들의 신뢰성과 일반화 가능성은 단일 기관 데이터에 대한 의존, 단일 임상 결과 측정에 대한 제한, 데이터 부족시 차별적인 성능의 감소로 인해 의문을 제기했다. 가장 중요한 점으로서, 효과적인 자원 활용에 중요한 질병 중증도 예측 모델에 대한 환자 분류 임계 값은 제안된 모델들의 실제 구현 가능성을 제한할 수 있음에도 불구하고 무시된 주제였다. 자원 제한을 고려한 기존 시뮬레이션 연구는 COVID-19의 맥락에서 기계 학습 모델 또는 환자 유입 시나리오를 함께 활용하지 않았다. 종합하면, 이러한 모델들의 단점들은 다수의 대규모 기관 데이터를 기계 학습 및 시뮬레이션 모델링과 같은 고급 예측 모델과 결합해야 할 필요성을 강조한다.

[0102] 본 출원인의 연구의 목적은 세 가지였다. 먼저, 본 출원인은 COVID-19 환자를 분류하기 위한 설명 특성이 있는 기존 예측 모델을 개발하는 것을 목표로 했다. 둘째, 본 출원인은 기존 예측 모델을 기반으로 재귀적 특성 감소(recursive feature elimination, RFE) 기술을 사용하여 제한된 데이터 가용성 하에서 효율적인 활용을 위해 변수 감소 모델을 개발했다. 마지막으로, 본 출원인은 제한된 의료 자원에서 효율적인 자원 활용을 위해 이산 이벤트 시뮬레이션(discrete event simulation, DES) 기술을 사용하여 적응형 분류 모델을 개발했다. 본 연구에서 개발 및 검증된 모델은 의료 시스템 용량을 유지하면서 환자 사망률을 최소화하기 위해 COVID-19 환자를 분류하는 데 큰 가치가 있을 것이다.

[0104] 도 5는 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델을 생성하기 위한 훈련 코호트 및 홀드-아웃 코호트의 설명적 특성을 도시한다.

[0106] 방법

[0107] 윤리적 진술(Ethical statement)

[0108] 본 발명의 연구는 병원윤리위원회 (2020-0883-001)와 질병관리청 (Korea disease control and prevention agency, KDCA) 역학조사분석위원회 (20201120_4a)의 승인을 받았다. 모든 연구 절차는 1946 년 헬싱키 선언과 2008 년 개정안을 준수했다.

[0110] 환자 코호트(Patient cohort)

[0111] 본 출원인은 2020 년 4 월까지 중증 급성 호흡기 증후군 코로나 바이러스에 대한 비인두 / 구강인두 면봉 또는 가래 검체를 사용하여 실시간 역전사 중합효소 연쇄반응 (polymerase chain reaction, PCR)으로 확진된 5628 명의 환자의 인구학적 데이터, 임상적 데이터, 실험실 데이터 및 질병 결과 기록을 후향적으로 검색했다. 환자의 데이터는 KDCA에서 수집하고 종합적으로 관리했다. 데이터베이스에는 격리 또는 입원에서 치료를 받고 퇴원한 환자와 COVID-19의 후유증으로 사망한 환자가 포함되었다. 격리 종료 기준은 최소 24 시간 이상 간격을 두고 두 번의 연속 음성 결과가 나오는 것 및 환자가 무증상인 것이었다. CSS 데이터가 누락된 환자는 제외되어 최종 개발 코호트는 5601 명 (99.5 %)이었다.

[0113] 공변량 및 결과 정의(Covariates and outcome definition)

[0114] 진단시 수집된 인구학적 데이터 (demographic data), 의학적 동반 질환 (medical comorbidities), 임상적 데이

터 (clinical data) 및 실험실 데이터 (laboratory data)를 모델 개발에 사용했다.

[0115] 인구학적 데이터 (demographic data)는 환자 나이 (patient age), 성별 (sex), 수축기 혈압 (systolic blood pressures), 이완기 혈압 (diastolic blood pressures), 심박수 (heart rate), 체온 (body temperature), 체질량 지수 (body mass index)의 7가지 데이터 유형을 포함했다.

[0116] 의학적 동반 질환 (medical comorbidities)은 고혈압 (hypertension), 당뇨병 (diabetes mellitus), 심부전 (heart failure), 심혈관 질환 (cardiovascular disease), 천식 (asthma), 만성 신장 질환 (chronic kidney disease), 만성 폐쇄성 폐 질환 (chronic obstructive pulmonary disease), 만성 간 질환 (chronic liver disease), 자가 면역 질환 (autoimmune disease), 치매 (dementia), 악성 종양 (malignancy), 임신 (pregnancy) 및 임신 주수 (pregnancy weeks)의 12가지 데이터 유형을 포함했다.

[0117] 임상적 데이터 (clinical data)는 발열 (fever), 기침 (cough), 객담 생성 (sputum production), 근육통 (myalgia), 피로 (fatigue), 인후통 (sore throat), 콧물 (rhinorrhea), 호흡 곤란 (shortness of breath), 구토 (vomiting), 설사 (diarrhea), 두통 (headache) 및 의식 변화 (altered consciousness)의 13가지 데이터 유형을 포함했다.

[0118] 실험실 데이터 (laboratory data)는 헤모글로빈 (hemoglobin), 헤마토크릿 (hematocrit), 백혈구 수 (white blood cell count), 림프구 (lymphocyte) 및 혈소판 수 (platelet count)의 5가지 데이터 유형을 포함했다.

[0119] 격리 또는 입원 중 최대 임상 중증도 점수 (clinical severity score, CSS)는 다음과 같이 분류되었다: 활동 제한 없음 (1), 활동 제한은 있지만 산소 요법은 없음 (2), 비강을 이용한 산소 요법 (3), 얼굴 마스크를 사용한 산소 요법 (4), 비 침습적 환기 (5), 침습적 환기 (6), 체외 막 산소화를 이용한 장기 지원 (7), 죽음 (8). 본 발명의 연구에서 CSS ≥ 5 의 정의는 임상 개선을 위한 세계보건기구 순서 척도 (WHO Ordinal Scale)의 정의와 비슷했다 (comparable).

[0121] 결과 (Results)

[0122] 환자 특성들 (Patient characteristics)

[0123] 도 5는 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델을 생성하기 위한 훈련 코호트 및 홀드-아웃 코호트의 설명적 특성을 제공한다. 총 5330 명 (95.2 %)의 환자는 임상 중증도 점수 (clinical severity score, CSS)가 6 미만으로서 비-중증 질환을 나타낸 반면, 271 명 (4.8 %)의 환자는 CSS가 6 이상으로서 중증 질환을 나타냈다.

[0125] 도 6은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델을 생성하기 위한 모델 별 성능 측정 결과를 도시한다.

[0127] 통계 분석 (Statistical analysis)

[0128] 모델 개발 (Model development)

[0129] 다변량 LR 및 XGBoost를 사용하여 가장 성능이 좋은 예측 모델을 선택했다. 본 발명의 모델은 5037 명 (89.9 %)의 환자 데이터를 사용하여 훈련 및 교차 검증되었으며 추가 검증을 위해 564 명 (11.1 %)의 홀드-아웃 코호트를 사용하여 재 검증되었다. LR 모델에서는 2674 명의 환자 데이터가 사용된 반면, XGBoost 모델에서는 모든 환자 데이터가 누락된 데이터를 처리하는 특성의 이점으로 사용되었다. 성능 지표는 과적합을 피하기 위해 10 배 교차 검증 (10-fold cross-validation)을 사용하여 계산되었다. 모델 개발은 R 프로그램 (R program)의 캐럿 (caret) 패키지를 사용하여 수행되었다.

[0131] 모델 성능 (Model performance)

[0132] 환자 중증도를 예측하기 위해 극한 기울기 부스팅 (extreme gradient boosting, XGBoost) 및 로지스틱 회귀 (logistic regression, LR)를 사용했다.

- [0133] 도 6을 참조하면, XGBoost 및 LR 모델의 교차 검증된 수신기 작동 특성 영역 (area under the receiver operating characteristics, AUROC)은 각각 0.97 (95 % 신뢰 구간 [confidence interval, CI] : 0.96-0.97) 및 0.94 (95 % CI : 0.91-0.96) 였다 ($p = 0.04$). XGBoost는 모든 성능 측정에서 LR을 능가했다. 본 출원인은 XGBoost가 4 개의 CSS 엔드포인트들에서 LR을 능가했으며 그 중 2 개가 통계적으로 유의하다는 점을 고려하여 XGBoost 모델을 기준 모델로 선택했다.
- [0135] **모델 비교 (Model comparison)**
- [0136] XGBoost 알고리즘의 하이퍼 파라미터는 10 배의 교차 검증 (10-fold cross-validation)과 함께 간단한 그리드 검색을 사용하여 AUROC를 최대화하도록 최적화되었다. 누락된 데이터를 대체하거나 특성을 선택하는 것과 같은 전처리 (preprocessing) 방법을 사용하지 않고 모델 성능을 비교했다.
- [0137] 정확도, AUROC, 감도, PPV 및 NPV는 R의 pROC 패키지를 사용하여 90 % 특이성에서 계산되었다. 성능 측정의 CI 는 2,000 개의 반복된 계층화된 부트 스트랩 (bootstrap) 방법을 사용하여 계산되었다. 모델 성능 차이에 대한 p 값도 앞서 언급한 부트 스트랩 방법을 사용하여 계산되었다.
- [0139] 도 7은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델에 대한 각 예측 변수의 샐플리 값 (shapley value, SHAP value)을 도시한다. 구체적으로, 도 7은 각각의 특성과 SHAP 값들 사이의 관계 (relationships between each feature and SHAP values)를 도시한다.
- [0140] 도 7은 각 점이 모델 훈련에 사용된 데이터에서 환자의 SHAP 값을 나타내는 요약 플롯이다. 모델에서 선택하지 않은 두 가지 특성 (임신 및 임신 주)을 제외하고 기준 모델을 맞추는 데 사용되는 모든 특성에 대해 점이 표시 된다. SHAP 값은 각 점의 절대 SHAP 값을 평균하여 계산된 y 축의 특성 중요도에 대한 순위 순서로 표시된다. x 축의 점 위치는 모델의 예측 출력에 미치는 영향을 보여준다. 보라색은 상대적으로 높은 특성 값을 나타내고 노란색은 상대적으로 낮은 특성 값을 나타낸다. 회색 점은 손실 값 (missing value)을 나타낸다.
- [0142] 도 8은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델에 대한 각 예측 변수 별 샐플리 값 (shapley value)을 도시한다. 구체적으로, 도 8은 각각의 특성과 SHAP 값들 사이의 관계 (relationships between each feature and SHAP values)를 도시한다.
- [0143] 도 8은 나이 (age), 림프구 (lymphocyte), 혈소판 (platelet), 체질량 지수 (body mass index), 헤마토크릿 (hematocrit), 숨가쁨 (shortness of breath), 성별 (sex), 체온 (body temperature), 심박수 (heart rate) 등 상위 9 개 주요 특징에 대한 SHAP 값들의 의존도를 도시한다. 각 그래프는 모델의 예측에 대한 각 특성의 영향을 나타낸다. x 축은 변수 값을 나타내고 y 축은 SHAP 값을 나타낸다. 변곡점은 모델의 예측에 대한 특징의 비선형 영향을 나타낸다.
- [0145] 도 9는 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델의 상위 10개 예측 변수에 대한 실제 중증 환자 특정 SHAP 플롯을 도시한다. 구체적으로, 도 9는 환자 특성 SHAP 플롯 (patient-specific SHAP plots)을 도시한다. 도 9에서 평균 SHAP 값에 따라 각 샘플 환자에 대해 순위가 매겨진 상위 10 개 특성은 모델의 예측 동작을 나타낸다.
- [0146] 도 9를 참조하면, 실제 중증 환자에 대한 SHAP 플롯은 나이 (age), 림프구 (lymphocyte), 숨가쁨 (shortness of breath)이 양의 방향으로 예측 출력에 상당한 영향을 미친다는 것을 보여준다. 그러나, 백혈구 (white blood cell, WBC) 수, 헤마토크릿 (hematocrit), 헤모글로빈 (hemoglobin), 체질량 지수 (body mass index)는 역으로 기여한다.
- [0148] 도 10은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델의 상위 10개 예측 변수에 대한 실제 비-중증 환자 특정 SHAP 플롯을 도시한다. 구체적으로, 도 10은 환자 특성 SHAP 플롯 (patient-specific SHAP plots)을 도시한다. 도 10에서 평균 SHAP 값에 따라 각 샘플 환자에 대해 순위가 매겨진 상위 10 개 특성은 모델의 예측

동작을 나타낸다.

[0149] 도 10을 참조하면, 실제 중증이 아닌 환자에 대한 SHAP 플롯은 나이 (age), 림프구 (lymphocyte), 혈소판 (platelet) 수 및 헤모글로빈 (hemoglobin)이 음의 방향으로 예측 출력에 상당한 영향을 미친다는 것을 나타낸다. 그러나, 수축기 혈압 (systolic blood pressure)은 역으로 기여한다.

[0151] 모델 해석 가능성 (Model interpretability)

[0152] 블랙박스 기계학습 모델을 해석하기 위하여 샐플리 추가 설명 (shapley additive explanations, SHAP)이 사용되었다. SHAP는 각 데이터의 샐플리 값 (shapley value)을 기반으로 출력에 얼마나 영향을 미치는지 계산하는 알고리즘이다. SHAP에 따르면 나이와 림프구 수는 $CSS \geq 6$ 의 질병 중증도에 대한 가장 중요한 위험 요소였다. 예측 변수의 강도와 방향은 도 7에 나와 있다. 나이 (age), 림프구 (lymphocyte), 혈소판 (platelet), 체질량 지수 (body mass index), 헤마토크릿 (hematocrit) 및 심박수 (heart rate)는 중증도 예측에 비선형적인 영향을 나타냈다. 도 8을 참조하면, 비선형성은 명확한 변곡점에서 추론할 수 있다. 도 9 및 도 10을 참조하면, 모델 출력에 대한 각 특성의 전체적인 영향 외에도 SHAP는 예측된 중증도에 대한 각 변수의 환자 특정 속성을 제공한다.

[0154] 제한된 데이터 가용성에서 예측 가능한 성능 (Predictive performance under limited data availability)

[0155] 이하, F+L+ 모델은 실험실 데이터가 있는 전체 모델 (full models with laboratory data)이다. F+L- 모델은 실험실 데이터가 없는 전체 모델 (full models without laboratory data)이다. F-L+ 모델은 실험실 데이터가 있는 감소된 모델 (reduced model with laboratory data)이다. F-L- 모델은 실험실 데이터가 없는 감소된 모델 (reduced model without laboratory data)이다.

[0156] 재귀적 특성 감소 (recursive feature elimination, RFE)는 기준 모델을 개선하기 위해 수행되었으며 제한된 데이터 가용성 하에서 성능이 검증되었다. AUROC 0.97 (95 % CI : 0.96-0.97)은 37 개 변수 (임상 변수 32 개와 실험실 변수 5 개)를 모두 포함하는 기준 모델에서 얻었다.

[0158] 도 11은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델 중 실험실 변수를 포함한 모델 (F+L+ model)의 AUROC (area under the receiver operating characteristic)를 도시한다. 구체적으로, 도 11은 RFE 적용시 모델 성능의 변화 (the change of model performance when applying RFE)를 도시한다. 도 11에서 재귀 특성 감소 (recursive feature elimination, RFE) 전체 과정의 결과는 AUROC (area under the receiver operating characteristic)의 관점에서 제시되었다. 각 포인트는 가장 낮은 순위의 특성이 감소된 후 모델의 성능을 나타낸다.

[0159] 도 11은 실험실 변수를 포함한 모델의 AUROC를 도시한다. RFE는 5 개의 실험실 변수를 포함하여 37 개의 변수로 구성된 F+L+ 세트에서 수행되었다. 첫 번째 반복에서 F+L+ 모델은 37 개의 변수에 대해 훈련된 반면, F-L+ 모델은 17 개의 변수를 사용하여 훈련되었다.

[0161] 도 12는 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델 중 실험실 변수를 제외한 모델 (F+L-)의 AUROC (area under the receiver operating characteristic)를 도시한다. 구체적으로, 도 12는 RFE 적용시 모델 성능의 변화 (the change of model performance when applying RFE)를 도시한다. 도 12에서 재귀 특성 감소 (recursive feature elimination, RFE) 전체 과정의 결과는 AUROC (area under the receiver operating characteristic)의 관점에서 제시되었다. 각 포인트는 가장 낮은 순위의 특성이 감소된 후 모델의 성능을 나타낸다.

[0162] 도 12는 실험실 변수를 제외한 모델의 AUROC를 도시한다. RFE는 5 개의 실험실 변수가 제외된 32 개의 변수로 구성된 F+L- 세트에 대해 수행되었다. 첫 번째 반복에서 F+L- 모델은 32 개의 변수에 대해 훈련된 반면 F-L- 모델은 11 개의 변수를 사용하여 훈련되었다.

- [0164] 도 13은 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델 중 실험실 변수를 포함한 모델 (F+L+ model) 및 실험실 변수를 제외한 모델 (F+L-)에 대하여 예측 변수의 개수에 따른 성능을 도시한다.
- [0165] 도 11 및 도 13을 참조하면, 성능 저하는 20 개의 변수가 감소될 때까지 통계적으로 미미하여 F-L+ 모델이 생성되었다. F+L+ 모델은 민감도와 특이도 모두 90 % 이상을 달성했다. 그러나, 상대적으로 낮은 양성 예측 값 (positive predicted value, PPV)은 31 %이다. 본 출원인은 중증도 분포의 불균형이 이 결과에 영향을 미칠 수 있다고 가정했다. 17 개의 변수로 구성된 F-L+ 모델은 90 %의 고정 특이도 수준에서 민감도 88 %, PPV 31 %를 달성했다. F-L+ 모델은 여전히 AUROC, 민감도, 특이성, PPV 및 음성 예측 값 (negative predicted value, NPV)과 관련하여 LR 모델을 능가했다.
- [0166] 도 12 및 도 13을 참조하면, 본 출원인은 32 개의 변수를 포함하는 F+L- 모델에서 0.95 (95 % CI : 0.94-0.96)의 AUROC를 얻었다. 성능 감소는 21개의 변수가 감소될 때까지 통계적으로 미미하여 F-L- 모델이 생성되었다.
- [0168] 도 14는 본 발명의 다양한 실시 예들에 따른 각각의 중증도 예측 모델 별 성능을 도시한다.
- [0169] F+L- 및 F-L- 모델은 각각 90 %의 고정 특이성 수준에서 민감도 84 % 및 81 %를 달성했다. 도 14는 정확도, PPV 및 NPV에 대한 자세한 비교를 요약한다. AUROC의 중요한 차이는 실험실 데이터가 있거나 또는 없는 전체 모델 (F+L+ 및 F+L-)과 실험실 데이터가 있거나 또는 없는 감소된 모델 (F-L+ 및 F-L-) 모두에서 관찰되었으며, 이는 실험실 데이터의 확실한 차별성을 의미한다 (모두 $p \leq 0.01$).
- [0171] 도 15는 본 발명의 다양한 실시 예들에 따른 각각의 중증도 예측 모델 별 과적합(over-fitting)을 방지하기 위한 검증 결과를 도시한다.
- [0172] 과적합(over-fitting)을 방지하기 위해 홀드-아웃 된 데이터 세트에 대해 네 가지 모델을 검증했다. F+L+ 및 F+L- 모델의 AUROC는 각각 0.958 (95 % CI : 0.924-0.991) 및 0.943 (95 % CI : 0.901-0.985)이었으며 교차 검증 결과와 통계적으로 유의한 차이가 없었다 (statistically indifferent) (각각 $p = 0.663$ 및 $p = 0.889$). F-L+ 및 F-L- 모델의 AUROC는 각각 0.949 (95 % CI : 0.906-0.990) 및 0.941 (95 % CI : 0.903-0.978)이었으며 교차 검증 결과와 통계적으로 무관했다 (각각 $p = 0.541$ 및 $p = 0.950$). 도 15는 교차 검증된 결과와 홀드-아웃 된 결과 간의 무관함은 모든 모델이 보이지 않는 데이터에 대한 일반화 가능성을 가짐을 보여준다.
- [0174] 도 16는 본 발명의 다양한 실시 예들에 따른 중증도 예측 모델에 대하여 재귀적 특성 감소 (recursive feature elimination, RFE)를 적용하기 위한 예측 변수의 순위를 도시한다.
- [0175] 도 14 및 도 16은 RFE의 각 단계에서 자세한 결과와 선택된 변수를 제공한다.
- [0177] **변수 감소 (Variable elimination)**
- [0178] RFE 기술은 많은 예측 변수가 감소되었을 때 모델 성능 유지 범위를 평가하는 데 사용되었다. RFE는 F+L+ 모델 (실험실 데이터가 있는 전체 모델)과 F+L- 모델 (실험실 데이터가 없는 전체 모델)에서 수행되었다. 그 결과 F-L+ 및 F-L- 모델은 변수 수가 가장 적을 때까지 도출되었으며, AUROC는 각 전체 모델과 비교할 때 통계적 차이를 보이지 않았다.
- [0179] SHAP를 사용하여 모델에 대한 중요도를 기준으로 로컬 정확도, 누락 및 일관성과 같은 바람직한 속성에 대한 순위를 매겼다. RFE를 반복할 때마다 가장 낮은 순위의 특성이 감소되고 모델이 재적합 (re-fit) 되었으며 성능은 10 겹 교차 검증을 사용하여 평가되었다. 축소된 모델은 AUROC의 차이가 통계적 유의성을 보이지 않는 반면 특징 수가 최소화된 지점에서 선택되었다.
- [0180] 4 개의 분류 모델은 과적합을 방지하기 위해 홀드-아웃 코호트로 재검증되었다. 특징 감소 및 특징 중요도 계산은 각각 R 프로그램 (R program)의 캐럿 (caret) 패키지 및 SHAPforXGBOOST 패키지를 사용하여 수행되었다.

[0182] **모델 해석 가능성 (Model interpretability)**

[0183] FHL+ 모델을 해석하기 위해 본 출원인은 블랙박스 기계학습 모델에 가시적인 해석성을 제공하는 SHAP를 사용했다. 요약 플롯과 종속성 플롯이 생성되었다. 특정 예측의 특징 SHAP 점수를 집계하여 환자 특정 플롯을 만들었다.

[0185] 도 17은 본 발명의 다양한 실시 예들에 따른 제한된 의료 자원 가용성 하에서 환자 사망률 시뮬레이션을 생성하기 위한 워크 플로우를 도시한다. 구체적으로, 도 17은 기계학습 기반 분류 시스템을 통해 COVID-19 환자를 의료 자원에 할당하는 방법을 보여주는 다이어그램이다.

[0187] **제한된 자원 가용성 하에서 최적의 환자 분류 (Optimal triage under limited resource availability)**

[0188] 환자 유입량과 의료 자원의 용량에 따라 조정할 수 있는 최적의 환자 분류 임계 값을 정의하기 위해 이산 이벤트 시뮬레이션 (discrete event simulation, DES)를 수행했다. 도 17은 DES를 수행함에 있어서 전체 시뮬레이션 워크 플로우를 도시한다.

[0189] 도 17의 실시 예에서 중증은 CSS가 6 이상인 경우를 의미한다. $P_{severity}$ 는 중증도 예측 모델에 의하여 환자의 중증도가 6 이상일 확률을 의미하며, 구체적으로, 입원 기간 중 환자의 CSS가 최대 6 이상으로 이르게 될 확률을 의미한다. 도 17의 실시 예에서 임계 값 (threshold)는 CSS가 6 이상에 이르게 될 확률인 $P_{severity}$ 가 얼마 이상일 때 중증으로 볼 것인지를 결정하는 기준이다.

[0190] 특정 임계 값 이상의 중증도 예측 확률을 가진 환자는 용량 가용성에 따라 중환자 실 (intensive care unit, ICU)로 우선 분류된다. 반면에, ICU가 사용할 수 없는 상태이거나 환자의 중증도 예측 확률이 특정 임계 값 미만인 경우, 환자는 일반 병동으로 분류된다.

[0191] I 유형 사망은 ICU에서 환자의 사망을 나타낸다. II 유형 사망은 ICU를 사용할 수 없는 상태로 인해 일반 병동으로 이송된 환자의 사망을 나타낸다. III 유형 사망은 중증도 예측 확률이 특정 임계 값 미만이어서 일반 병동으로 이송된 환자의 사망을 나타낸다.

[0192] 시뮬레이션을 사용하여 사망률 = $(\text{총 사망자 수} / \text{총 환자 수}) = ((\text{I 유형 사망자 수} + \text{II 유형 사망자 수} + \text{III 유형 사망자 수}) / \text{총 환자 수})$ 을 최소화 할 수 있다.

[0194] 도 18은 본 발명의 다양한 실시 예들에 따른 환자 유입 시뮬레이션을 생성하기 위한 과거 환자 유입 시나리오를 도시한다. 구체적으로, 도 18은 환자 유입 (patient influx)을 도시한다. 도 18은 한국에서 COVID-19의 역사적 환자 유입에 대한 감염병 발생률 곡선을 나타낸다.

[0195] 도 18에서 H1, H2, H3, H4는 한국에서 COVID-19의 역사적 환자 유입 시나리오를 도시한다. 과거 역사적 환자 유입 시나리오 H1, H2, H3 및 H4에 대해 사망률은 각각 0.1, 0.01, 0.04 및 0.24의 임계 값에서 최소화되었다.

[0197] **임계 값 최적화 (Optimization of threshold)**

[0198] 이산 이벤트 시뮬레이션 및 환자 유입 생성 (Discrete events simulation (DES) and patient influx generation).

[0199] 시뮬레이션을 위해 simmer R 패키지를 사용하여 제한된 의료 자원에서 $(\text{총 사망자 수} / \text{총 환자 수})$ 로 계산된 사망률을 최소화하는 최적의 임계 값을 찾기 위해 DES를 수행했다.

[0200] 도 18을 참조하면, 2020 년 2 월부터 2021 년 2 월 사이에 관찰된 4 가지 COVID-19 역사적 감염병 환자 유입 시나리오 (H1, H2, H3, H4)를 사용하여 시뮬레이션을 수행했다.

[0202] 도 19는 본 발명의 다양한 실시 예들에 따른 기초 감염재생산 지수 (basic reproduction rate) (R_0)에 기반한

가상의 환자 유입 시나리오를 도시한다. 구체적으로, 도 19는 환자 유입 (patient influx)을 도시한다.

- [0203] 도 19는 SIR 시뮬레이션을 이용한 가상의 환자 유입 시나리오에 따른 곡선을 도시한다. 도 19는 기초 감염재생산 지수 (basic reproduction rate) (R_0)에 기반한 가상 환자 유입과 그 설명 통계를 나타낸다. R_0 는 감염병이 사람 간 전파되는 정도를 수치로 나타낸다. R_0 가 1보다 크면 전염병이 감염자 1명에게서 다른 사람 1명 이상으로 전파된다는 의미다. R_0 값이 사스는 4이며, 메르스는 0.4~0.9이었다.
- [0204] 도 19를 참조하면, 가상의 환자 유입 시나리오는 전체 인구가 감염가능자 (susceptible), 감염자 (infected), 회복자 (recovered)의 세 집단으로 구성된 SIR 모델 (감염가능자 (susceptible) - 감염자 (infected) - 회복자 (recovered) model)을 사용하여 생성되었다.
- [0205] SIR 모델은 감염병 모델링을 위한 수학 모형이다. SIR 모델에서는 숙주를 감수성군(S), 감염군(I), 회복군(R)으로 분류한다. 즉, S는 'Susceptible number'로 감염 가능한 사람의 수를 나타내고, I는 'Infectious number'로 감염된 환자의 수를 나타내며, R은 'Recovered(or immune) number'로 질병으로부터 회복된 환자 (또는, 면역을 가진 자)의 수를 나타낸다. 이에 따라, 질병은 $S \rightarrow I \rightarrow R$ 의 과정을 통하여 감염이 확산되었다가 회복되는 형태를 나타내게 된다. SIR 모델의 미분 방정식을 표현하면 아래 [수학식 2]와 같다.
- [0206] [수학식 2]
- [0207]
$$dS/dt = -\beta \cdot S \cdot I/N$$
- [0208]
$$dI/dt = \beta \cdot S \cdot I/N - \gamma \cdot I$$
- [0209]
$$dR/dt = \gamma \cdot I$$
- [0210] [수학식 2]에서 β 는 감염률이고, γ 는 회복률이고, S는 감염 가능한 사람의 수이고, I는 감염된 사람의 수이고, R은 회복된 사람의 수이고, t는 시간이고, $N=S(t)+I(t)+R(t)$ 이고 고정된 총 인구 수를 나타낸다.
- [0211] $N = S(t) + I(t) + R(t)$ 로 계산된 전체 인구는 한국에서 관찰된 최대 역사적 환자 유입인 58,654 (H4)를 고려할 때 60,000으로 고정되었다. $I(0)$ 과 $R(0)$ 은 각각 6과 0으로 고정되었다. 한국에서는 평균 회복 시간이 20.1일이었기 때문에 회복률 γ 은 0.05로 설정되었다. 감염률 β 은 서로 다른 R_0 (기초 감염재생산 지수) 값에 따라서 0.75에서 5 사이의 범위에서 환자 유입을 생성했다. 매일 새로 확진된 환자의 수는 SIR 모델링에서 얻었다. 일일 확진 환자는 과거 역사적 환자 유입과 동일한 방법으로 변환되어 일일 도착 시간이 균일하게 적용되었다.
- [0213] 도 20은 본 발명의 다양한 실시 예들에 따른 중증 환자 유입 시뮬레이션을 생성하기 위한 중증 및 중증이 아닌 환자의 예측 확률 분포와 생성된 예측 확률 분포를 도시한다. 구체적으로, 도 20은 out-of-fold 샘플에서 환자의 예측 확률 분포 그래프를 도시한다.
- [0214] 도 20은 실제 데이터에서 관찰되는 환자의 분포 (본 연구의 경우, 5천명)를 바탕으로 가상의 시뮬레이션을 위한 환자의 분포 (6만명)를 만들었을 때, 상기 두 분포가 다르지 않음을 보이기 위한 그래프를 나타낸다. 중증인 (severe) 환자가 전체의 4.8%이고 분포 양상이 다르기 때문에 각 분포도의 생성은 개별적으로 진행되었다. 적분 값이 1인 커널 함수 (Kernel function)을 이용하여 확률 밀도 (density)를 추정하였다. 관측된 데이터들의 분포로부터 변수의 분포 특성을 추정하였다. 도 20의 그래프에 대하여, y축은 확률 밀도 (density)를 나타내고, x축은 logits (log + odds)로서 중증인 (severe) 상태를 P로 놓았을 때 $\log(P / (1-P))$ 에 해당한다.
- [0216] **확률 생성(Probability generation)**
- [0217] 환자 중증도 확률의 경험적 확률 분포를 생성하기 위해 10 배의 교차 검증 (10-fold cross-validation)의 out-of-fold 예측 결과를 집계했다. 제한된 데이터 및 진단 도구에서 높은 성능과 잠재적인 사용을 고려하여 F+L-모델의 결과를 사용했다. 가우시안 커널 밀도 추정과 선형 보간을 사용하여 근사된 경험적 확률 분포 함수에 대해 역변환 샘플링을 수행했다. 커널 밀도 대역폭은 이전에 제안된 방법을 사용하여 선택되었다. 이 과정은 중증 및 중증이 아닌 환자에 대해 별도로 수행되었으며, 표본 확률은 중증도 데이터의 유행률을 유지하면서 생성된 환자 유입과 무작위로 일치되었다.
- [0218] 도 20은 out-of-fold 샘플에서 중증 및 중증이 아닌 환자의 예측 확률 분포와 생성된 예측 확률 분포를 나타낸다.

[0220] **시뮬레이션 시나리오 및 설정 (Simulation scenario and settings)**

[0221] 임계 값 이상의 확률을 가진 환자는 ICU로 안내되며 환자는 ICU의 수용 능력에 따라 입원한다. 거부된 환자는 임계 값 미만의 확률로 환자와 함께 일반 병동으로 안내된다. 과거 보고된 연구 결과에 따라서, ICU에 있는 동안 실제 중증 환자가 사망할 확률은 0.507로 설정되었고 ICU 외부에서 사망할 확률은 0.990으로 설정되었다. 본 출원인은 중증 환자가 아닌 환자는 ICU 입원과 관계없이 생존할 것이라고 가정했다. 환자 사망은 세 가지 유형으로 분류되었다. “중증”으로 분류되고 ICU로 이송되었으나 사망하는 유형 (I 유형); “중증”으로 분류되었으나 ICU 이용 불가로 인해 중증 환자가 사망하는 자원 의존적 사망의 유형 (II 유형); 및 중증 환자가 "비 중증"으로 잘못 분류되고 일반 병동으로 이송된 후 사망하는 환자 분류 임계 값 의존적 사망의 유형 (III 유형).

[0222] ICU의 최대 용량은 음압 하에서 격리 침대 수를 기준으로 504 개 침대로 설정되었다. 체류 기간 분포를 추정하기 위해 이전에 제안한 γ 분포를 사용했다. 중환자 실 환자의 평균 입원 기간은 γ 분포로 계산되었으며, 사망한 환자들의 경우, 형상 매개 변수 = 1.5488, 비율 매개 변수 = 0.1331이고, 중환자실 국가감사연구센터 (intensive care national audit and research center, ICNARC)에 제시된 사분위수 범위 및 근사 중앙 값까지 생존한 사람들의 경우, 형상 매개 변수 = 0.8904, 비율 매개 변수 = 0.0477 이었다. 입원 기간, 사망 및 생성된 중증도 확률의 무작위성에서 발생할 수 있는 이상치(outliers)에 대한 견고성을 보장하기 위해 각 환자 유입에 대해 서로 다른 시드로 시뮬레이션을 20 회 반복했다.

[0224] 도 21은 본 발명의 다양한 실시 예들에 따른 과거 역사적 환자 유입 시나리오에서 중증도 판단의 임계 확률에 따른 사망률의 그래프를 도시한다. 구체적으로, 도 21은 다양한 시나리오에서 각 임계 값의 사망률. 검은 색 점은 최적화된 사망률과 각 환자 유입에 대한 최적의 임계 값을 나타낸다. 도 21은 임계 값에 따른 역사적 환자 유입의 사망률을 도시한다.

[0225] 도 21을 참조하면, 사망률은 임계 값에 따라 불룩한 모양을 보인다.

[0227] 도 22는 본 발명의 다양한 실시 예들에 따른 과거 역사적 환자 유입 시나리오에서 중증도 판단의 임계 확률에 따른 사망 하위 유형 별로 분해된 사망률의 그래프를 도시한다.

[0228] 구체적으로, 도 22는 각 임계 값에서 사망 하위 유형 별로 분해된 사망률 (mortality rate decomposed by death subtype at each threshold)를 도시한다. x 축은 임계 값을 나타내고, y 축은 각 임계 값에서 ((사망 하위 유형 수) / (총 환자 수))로 계산 된 총 환자에 대한 각 사망 하위 유형의 누적 비율을 나타낸다. 도 22는 역사적 환자 유입에 따른 각 임계 값에서 사망 하위 유형 별로 분해된 사망률을 도시한다.

[0229] 도 22를 참조하면, 사망률과 임계 값의 관계는 세 가지 유형의 사망 사이의 상충 관계로 설명할 수 있다. 본 출원인은 I 유형 사망 비율을 최대화하면 중환자 실 (intensive care unit, ICU) 내에서 사망할 확률 (I 유형 사망률 = 50.7 %)이 일반 병동 (II 유형 및 III 유형 사망률 = 99.0 %)보다 낮기 때문에 최소 사망률로 이어진다는 사실을 발견하였다.

[0231] 도 23은 본 발명의 다양한 실시 예들에 따른 과거 역사적 환자 유입 시나리오 별 최적 임계 확률, 최적화된 사망률, 감소된 사망률을 도시한다. 구체적으로, 도 23은 환자 분류 시뮬레이션의 최적화된 결과 (optimized results of patient triage simulation)를 도시한다. 도 23은 역사적 환자 유입의 시뮬레이션 결과를 도시한다.

[0232] 무제한 ICU 용량에서 임계 값을 낮추면 ICU에 입원한 실제 중증 사례를 포함하여 더 많은 환자가 발생하고 I 유형 사망이 증가한다. 그러나, 제한된 ICU 용량에서 임계 값이 너무 낮으면 비-중증 환자가 잘못 분류되어 ICU 용량이 고갈될 수 있다. 결과적으로, 실제 중증 환자에 대한 제한된 용량은 상대적으로 I 유형 사망을 감소시키고 II 유형 사망을 증가시킨다. 반면, 임계 값이 너무 높으면 ICU 입원 거부율이 높아져 I 유형 사망이 감소하고 III 유형 사망이 증가할 수 있다.

[0233] 환자 유입 내 사망률과 환자 분류 임계 값 간의 연관성은 과거 역사적 환자 유입 데이터와의 분석을 통해 추론

할 수 있지만 일반적인 결론을 내리는 것은 비현실적이다.

- [0234] 예를 들어, 도 23을 참조하면, H2, H3 및 H4를 비교할 때 최적 임계 값 및 최적화된 사망률의 상승 추세를 찾을 수 있으며, 여기서, 환자 유입량의 각각 증가는 분명하다. 그러나, H1과 H3 또는 H4를 비교하는 것은 기간, 최대 일일 환자, 누적 환자와 같은 다차원적 특성의 차이로 인해 추론하기 어렵다. 결과를 더욱 뒷받침하기 위해 다양한 기초 감염재생산 지수 (basic reproduction rate) (R_0)을 가진 SIR 모델(감염가능자 (susceptible) - 감염자 (infected) - 회복자 (recovered) model)을 사용하여 생성된 환자 흐름을 사용하여 추가 시뮬레이션을 수행했다.
- [0236] 도 24은 본 발명의 다양한 실시 예들에 따른 기초 감염재생산 지수 (basic reproduction rate) (R_0)에 기반한 가상의 환자 유입 시나리오에서 중증도 판단의 임계 확률에 따른 사망률의 그래프를 도시한다. 구체적으로, 도 24은 다양한 시나리오에서 각 임계 값의 사망률. 검은 색 점은 최적화된 사망률과 각 환자 유입에 대한 최적의 임계 값을 나타낸다. 도 24은 임계 값에 따른 가상의 환자 유입의 사망률을 도시한다.
- [0237] 도 24를 참조하면, 가상 환자 유입이 있는 DES는 최적 임계 값이 0.02에서 0.66 범위였고, 각각의 최소화된 사망률은 0.017 (1.7 %)에서 0.042 (4.2 %) 범위였다. 각 R_0 에 대한 최적 임계 값 및 최소화된 사망률은 R_0 의 더 큰 값이 더 높은 최적 임계 값 및 사망률을 갖는 경향이 있음을 보여주었다.
- [0239] 도 25는 본 발명의 다양한 실시 예들에 따른 기초 감염재생산 지수 (basic reproduction rate) (R_0)에 기반한 가상의 환자 유입 시나리오에서 중증도 판단의 임계 확률에 따른 사망 하위 유형 별로 분해된 사망률의 그래프를 도시한다.
- [0240] 구체적으로, 도 25는 각 임계 값에서 사망 하위 유형 별로 분해된 사망률 (mortality rate decomposed by death subtype at each threshold)를 도시한다. x 축은 임계 값을 나타내고, y 축은 각 임계 값에서 (사망 하위 유형 수) / (총 환자 수)로 계산된 총 환자에 대한 각 사망 하위 유형의 누적 비율을 나타낸다. 도 25는 가상의 환자 유입에 따른 각 임계 값에서 사망 하위 유형 별로 분해된 사망률을 도시한다.
- [0241] 본 출원인은 각 R_0 에 대한 최적 임계 값 및 최소화된 사망률은 R_0 의 더 큰 값이 더 높은 최적 임계 값 및 사망률을 갖는 경향이 실제 수용 능력을 초과하는 ICU 수요 증가와 그로 인한 ICU 외부 사망률 증가 때문이라고 가정했다.
- [0242] 도 25를 참조하면, II 유형 및 유형 III 사망은 R_0 이 1.5 일 때 최적 임계 값에서 사망의 33.7 %를 차지한 반면, R_0 이 10 일 때 95.2 %를 차지했다. R_0 값이 높고 ICU 요구가 증가하는 상황에서 중증도가 낮은 환자를 거부해도 (예를 들어, 높은 환자 분류 임계 값 설정) 여전히 용량이 고갈되었다. 따라서, 임계 값을 조정하면 사망률을 크게 낮추지 않고 임계 값 종속 거부 횟수와 용량 종속 거부 횟수 간에 균형을 맞출 수 있다. 그러나, 이러한 상황에서 환자 분류 임계 값을 증가시키면 ICU 수용 환자의 정확도가 높아지면서 ICU 용량의 최대 활용도를 유지하여 사망률이 낮아진다.
- [0244] 도 26은 본 발명의 다양한 실시 예들에 따른 기초 감염재생산 지수 (basic reproduction rate) (R_0)에 기반한 가상의 환자 유입 시나리오 별 최적 임계 확률, 최적화된 사망률, 감소된 사망률을 도시한다.
- [0245] 구체적으로, 도 26은 환자 분류 시뮬레이션의 최적화된 결과 (optimized results of patient triage simulation)를 도시한다. 도 26은 가상의 환자 유입의 시뮬레이션 결과를 도시한다.
- [0246] 최적의 환자 분류 임계 값의 이점은 F-Lt 모델의 경우 0.013 인 기존 유텐 지수 (J-지수) (Youden Index (J-index))와 비교할 때 나타난다. J-지수를 사용한 사망률과 모든 유입에서 본 발명의 모델의 최소화된 사망률 간의 차이는 역사적 유입에서 4.7 % ~ 29.8 %, 가상 유입에서 6.1 % ~ 18.1 % 범위로서 현저하게 컸다. R_0 값이 작을 때 최적 임계 값과 J-지수 간의 차이는 작았다.
- [0247] 도 26을 참조하면, R_0 값이 지나치게 높은 환자 유입 시나리오 ($R_0 = 10$)에서 자원 용량을 초과하는 급증으로 인해 ICU 외부의 사망 (II 유형 및 III 유형 사망)이 극도로 지배적이 되어 (도 25에서, $R_0 = 10$), ICU 치료가 사망률에 미치는 영향을 제한했다.

[0249] 도 27은 본 발명의 다양한 실시 예들에 따른 각각의 과거 역사적 환자 유입 시나리오 및 기초 감염재생산 지수 (basic reproduction rate) (R_0)에 기반한 가상의 환자 유입 시나리오 별 최적 임계 확률, 최적화된 사망률, 감소된 사망률을 도시한다.

[0250] 도 26을 참조하면, R_0 값이 지나치게 높은 환자 유입 시나리오 ($R_0 = 10$)에서 자원 용량을 초과하는 급증으로 인해 ICU 외부의 사망 (II 유형 및 III 유형 사망)이 극도로 지배적이 되어 (도 25에서, $R_0 = 10$), ICU 치료가 사망률에 미치는 영향을 제한했다.

[0251] 그럼에도 불구하고, 도 27을 참조하면, 나머지 환자 유입 중 사망률이 크게 감소한 것은 자원 제약 하에서 환자 분류에 대한 임계 값의 최적화가 갖는 상당한 이점을 보여주었다.

[0253] 토론

[0254] XGBoost 알고리즘을 사용하여 COVID-19 진단시 환자의 정보를 캡처하여 SHAP를 통합하는 고성능 예측 모델을 통해 질병 중증도 결과에 영향을 미치는 기여 변수의 환자 특정 해석 가능성을 제공하였다. 그런 다음, RFE를 사용하여 제한된 데이터 가용성에서 모델의 성능이 지속 가능한지 확인했다. 마지막으로, 제한된 의료 자원에서 사망률을 최소화하기 위해 DES를 사용하는 최적의 분류 임계 값을 제안했다. 종합하면, 이러한 결과는 COVID-19의 향후 관리가 최대 환자 치료 효과를 보장하기 위해 의료 분야 및 헬스케어 분야 간에 통합될 수 있다는 잠재성을 보여주었다.

[0255] F+L+ 모델의 독특한 특징은 교차 검증 및 홀드-아웃 설정 모두에서 0.97을 초과하는 AUROC의 높은 판별력이었다. CALL, COVID-GRAM, ABC-GOALS 및 4C 사망률 점수 (Mortality Score)를 포함한 COVID-19 환자의 임상 악화에 대한 이전 예측 모델은 0.77에서 0.91까지 다양한 예측 정확도를 보고했다. 모든 모델에는 예측 정확도를 유지하기 위해 실험실 데이터 (laboratory data), 주변 산소 포화도 (peripheral oxygen saturation) 또는 방사선 학적 결과 (radiographic findings)를 포함한 특정 진단 데이터가 필요했다. 또한, 데이터가 부분적으로 없을 때 모델 성능이 어느 정도까지 유지되는지는 질병 중증도 예측 범위에서 아직 엄격하게 연구되지 않았다. 본 발명의 F+L+ 모델이 실제로 효과적으로 구현되도록 본 출원인은 감소된 모델(reduced model)이 임상 및 실험실 데이터가 부분적으로 없는 경우에도 적절한 식별력을 유지함을 확인했다. 감소된 모델의 장점은 보이지 않는 데이터에 대한 일반화 가능성뿐만 아니라 제한된 데이터 환경에서의 적용 가능성도 포함한다. 시기 적절한 환자 분류는 COVID-19의 주요 관심사이므로 감소된 모델은 F+L+ 모델이 환자의 위험을 보다 정확하게 계층화할 때까지 기다리는 동안 병원 도착시 조기 환자 분류에 활용할 수 있다. 감소된 모델은 고위험 환자를 위한 기계식 인공 호흡기 및 ICU와 같은 제한된 자원을 효율적으로 할당하기 위한 자원 제약 환경에서 특히 유용할 수 있다. 또한, COVID-19 환자에서 바이러스성 폐렴의 급성 악화를 감안할 때, 본 발명의 모델은 제한된 변수로 단기간에 입원 환자를 재평가하는 데 사용할 수 있으므로 임상 증상이 악화될 가능성이 있는 환자를 조기에 식별할 수 있다.

[0256] 본 발명의 모델의 주목할 특징은 SHAP 값을 사용하여 질병 악화에 대한 환자 특정 기여 요인과 기여 가중치를 구별하는 능력이다. 프레젠테이션에서 각 환자의 위험 요인에 대한 해석 가능성은 임상 의사에 대하여 개별화된 치료 계획을 안내할 수 있다. COVID-19 치료에 대한 현재 지침은 질병 개체에 대한 제한된 통찰력 하에서 평균 위험 환자를 기반으로 권장 사항을 제공한다. 이러한 권장 사항은 모든 환자에게 하나의 크기로 모든 환자에게 적용할 수 있는 접근 방식을 제공하며, 이는 복잡하거나 비정형적인 질병이 있는 환자에게는 문제가 될 수 있다. 이와 관련하여, 본 발명의 모델은 임의의 환자 위험 그룹화의 필요성을 제거하고 개별 위험 계층화를 기반으로 생존을 극대화하는 데 유용할 수 있다. 또한, 본 발명의 모델은 질병 악화의 불리한 위험 요소를 조기에 식별하기 위한 알림 시스템으로 사용되는 코딩 알고리즘을 활용하는 모든 전자 의료 기록 시스템에 통합될 수 있다.

[0257] 본 발명의 모델의 유효성은 SHAP를 통한 모델 해석 결과와 이전에 보고된 COVID-19 중증도의 예후 간의 높은 일관성으로 뒷받침된다. 본 출원인은 노년층이 림프 감소증과 혈소판 감소증에 이어 질병 악화에 대한 가장 높은 새플리 값 (shapley value)을 나타냈다는 점에 주목했다. 본 출원인은 노년층의 많은 나이가 호흡기 증후군 환자의 부작용 및 사망률과 관련된 특성 저하 및 허약함을 포함하여 노년층의 관련 특징과 상호 작용할 것이라고 가정했다. 본 출원인은 또한 림프구 감소증이 COVID-19에서 중요한 역할을 하며 잠재적인 치료 표적이 될 수 있다는 것을 추가한다. 림프구 감소증은 손상된 폐포 상피 세포에 의해 유도된 림프구의 침윤에 의해 발생하며 COVID-19 환자에서 흔히 관찰되었다. 이전 연구와 마찬가지로 혈소판 감소증도 질병의 부작용과 관련이 있었다.

폐가 성숙한 거핵 세포로부터 혈소판이 방출되는 부위이기 때문에 폐 모세 혈관 층의 감소 또는 형태학적 교대가 병리학적 혈소판 조각 모음을 발휘할 수 있다고 제안되었다. COVID-19 환자에서 림프구 감소증과 혈소판 감소증의 공통적인 공존에 근거한 또 다른 이론은 코로나 바이러스가 골수에 직접 침투하여 비정상적인 조혈 또는 혈액 세포에 대한 자가 면역 반응을 유발할 수 있다는 것이다. 본 발명의 예측 모델은 이러한 혈액학적 위기 이전에 코로나 바이러스 감염의 초기 식별이 더 나은 예후를 위해 필요하다는 개념을 뒷받침했다.

[0258] 본 출원인의 지식에 의하면, 자원 가용성을 설명하는 임계 값 수정을 사용하여 최적의 분류를 제공하기 위한 명시적 솔루션을 제안하는 COVID-19 중증도 예측 모델에 관한 기존 연구는 없다. DES는 개인, 인구 및 환경 간의 복잡한 행동과 상호 작용을 복제한다. 따라서, 자원 제약 하에서 사망률을 최소화하기 위해 더 나은 임상 결정을 내리기 위하여 널리 사용된다. 본 발명의 실시 예에 따르면, 본 출원인은 의료 인프라 중단과 관련된 COVID-19의 특징을 고려하여 F-L+ 모델로 DES를 수행하여 다양한 환자 유입 시나리오 및 동시 자원 가용성에서 적응적으로 활용할 수 있는 최적 임계 값을 찾았다. 실제로 이러한 도구는 자원의 잘못된 할당으로 인한 이환율과 사망률을 최소화하는 데 사용될 수 있다. 본 발명의 시뮬레이션은 과거 역사적 환자 유입 및 생성된 가상의 환자 유입을 모두 사용하는 최적의 임계 값이 각 특정 환자 유입 시나리오에 대한 사망률을 최소화할 수 있음을 보여주었다. COVID-19에 DES를 사용하는 연구는 나이를 유일한 결정 요인으로 사용하여 임계 값을 최적화하는 것이 사망률 감소에 제한적이거나 부정적인 영향을 미친다고 보고했다. 이러한 한계를 극복하기 위해, 본 발명은 최적의 예측 및 시뮬레이션 모델을 제공하기 위해 기계 학습 모델에서 얻은 중요한 변수를 활용했다. 본 출원인의 가설은 예상되는 환자 유입량에 따른 본 발명의 최적화된 임계 값 및 유덴의 J 통계 지수 (Youden's J-statistic) 사이의 사망률의 상당한 차이로 뒷받침된다. 이러한 관찰은 환자 분류 임계 값을 조정하여 사망률을 크게 줄일 수 있는 모델의 잠재력을 뒷받침한다.

[0259] 본 발명의 주목할 의미는 최적의 환자 분류의 효과가 기초 감염재생산 지수 (basic reproduction rate) (R_0)를 억제하려는 노력과 결합될 때 극대화된다는 것이다. 감염병이 의료 인프라에 미치는 영향을 완화하려는 기존 연구는 다양한 비-약물적 개입 (non-pharmacological intervention, NPI)을 통해 감염률을 줄이는 데 중점을 둔다. 그러나, 이러한 개입은 부정적인 경제적 영향과 같은 부작용이 있는 것으로 알려져 있으며, 보다 극단적인 조치가 시행될수록 악화될 수 있다. 바람직하지 않은 영향을 완화하고 제한된 의료 자원 하에서 동시에 사망률을 최소화하기 위해 최적의 환자 분류를 비-극단적인 비-약물적 개입과 함께 보완 조치로 활용할 수 있다.

[0260] 본 발명의 연구의 한계는 본 발명의 모델을 개발하기 위해 아시아 민족의 단일 국가 코호트를 통합한 것이었으며, 이러한 한계는 본 발명의 모델의 일반화 가능성을 제한할 수 있다. 유사한 차별적 성과를 보장하기 위해 다 민족 집단을 사용한 외부 검증이 필요하다. 하지만, 본 발명의 모델의 견고성을 보장하기 위해 홀드-아웃 코호트 (hold-out cohort)를 사용하여 추가 확인과 함께 10 배의 교차 검증 (10-fold cross-validation)을 구현했다. 본 발명의 연구의 또 다른 한계는 고유한 한계가 있는 시뮬레이션을 사용하여 환자 분류 임계 값을 평가했다는 것이다. 시뮬레이션은 구체적인 답을 얻지 못하며 모든 상황을 평가할 수도 없다. 또한, 본 발명의 연구의 범위를 벗어난 주제로서 향후의 환자 유입을 예측하는 것은 자원 용량의 포화 전에 최적의 임계 값을 찾기 위하여 반드시 필요하다.

[0261] 본 발명의 실시 예를 통하여, 제한된 데이터 및 자원 가용성에서도 COVID-19 환자 분류의 효율성을 향상시킬 수 있는 능숙한 수단을 제공하기 위해 설명 특성이 있는 강력한 예측 모델을 개발하고 검증했다. 본 발명의 모델은 코딩 알고리즘을 사용하는 모든 전자 의료 기록 시스템에 통합될 수 있기 때문에 널리 사용될 가능성이 있다. 또한, 환자 분류를 위한 최적의 임계 값을 설정하기 위해 동시 의료 자원 용량에 따라 환자 유입량을 활용하는 새로운 적응형 분류 모델을 제안했다. 본 발명의 연구는 COVID-19의 향후 효율적인 관리가 최대 환자 치료 효과를 보장하기 위해 의료 전문 지식 및 헬스케어 전문 지식을 통합해야 함을 의미한다.

[0263] 코드 가용성 (Code availability)

[0264] 모델 개발 및 평가에 사용되는 코드는 <https://github.com/minkim88/Optimal-Triage-COVID-19>에서 확인할 수 있다.

[0266] 본 연구는 한국보건산업진흥원 감염병 방역기술개발 사업 (HW20C2103)을 통해 지원되었다.

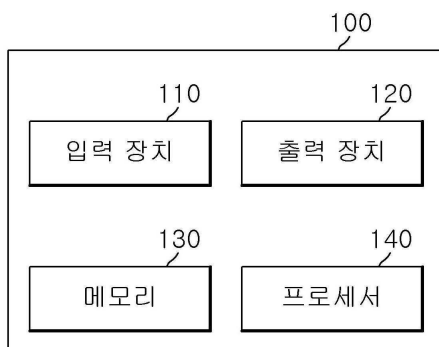
- [0268] 하드웨어를 이용하여 본 발명의 실시 예를 구현하는 경우에는, 본 발명을 수행하도록 구성된 ASICs(application specific integrated circuits) 또는 DSPs(digital signal processors), DSPDs(digital signal processing devices), PLDs(programmable logic devices), FPGAs(field programmable gate arrays) 등이 본 발명의 프로세서에 구비될 수 있다.
- [0269] 한편, 상술한 방법은, 컴퓨터에서 실행될 수 있는 프로그램으로 작성 가능하고, 컴퓨터 판독 가능 매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다. 또한, 상술한 방법에서 사용된 데이터의 구조는 컴퓨터 판독 가능한 저장 매체에 여러 수단을 통하여 기록될 수 있다. 본 발명의 다양한 방법들을 수행하기 위한 실행 가능한 컴퓨터 코드를 포함하는 저장 디바이스를 설명하기 위해 사용될 수 있는 프로그램 저장 디바이스들은, 반송파(carrier waves)나 신호들과 같이 일시적인 대상들은 포함하는 것으로 이해되지 않아야 한다. 상기 컴퓨터 판독 가능한 저장 매체는 마그네틱 저장매체(예를 들면, 롬, 플로피 디스크, 하드 디스크 등), 광학적 판독 매체(예를 들면, 시디롬, DVD 등)와 같은 저장 매체를 포함한다.
- [0270] 이상에서 설명된 실시 예들은 본 발명의 구성요소들과 특징들이 소정 형태로 결합된 것들이다. 각 구성요소 또는 특징은 별도의 명시적 언급이 없는 한 선택적인 것으로 고려되어야 한다. 각 구성요소 또는 특징은 다른 구성요소나 특징과 결합되지 않은 형태로 실시될 수 있다. 또한, 일부 구성요소들 및/또는 특징들을 결합하여 본 발명의 실시 예를 구성하는 것도 가능하다. 발명의 실시 예들에서 설명되는 동작들의 순서는 변경될 수 있다. 어느 실시 예의 일부 구성이나 특징은 다른 실시 예에 포함될 수 있고, 또는 다른 실시 예의 대응하는 구성 또는 특징과 교체될 수 있다. 특허청구범위에서 명시적인 인용 관계가 있지 않은 청구항들을 결합하여 실시 예를 구성하거나 출원 후의 보정에 의해 새로운 청구항으로 포함시킬 수 있음은 자명하다.
- [0271] 본 발명이 본 발명의 기술적 사상 및 본질적인 특징을 벗어나지 않고 다른 형태로 구체화될 수 있음은 본 발명이 속한 분야 통상의 기술자에게 명백할 것이다. 따라서, 상기 실시 예는 제한적인 것이 아니라 예시적인 모든 관점에서 고려되어야 한다. 본 발명의 권리범위는 첨부된 청구항의 합리적 해석 및 본 발명의 균등한 범위 내 가능한 모든 변화에 의하여 결정되어야 한다.

부호의 설명

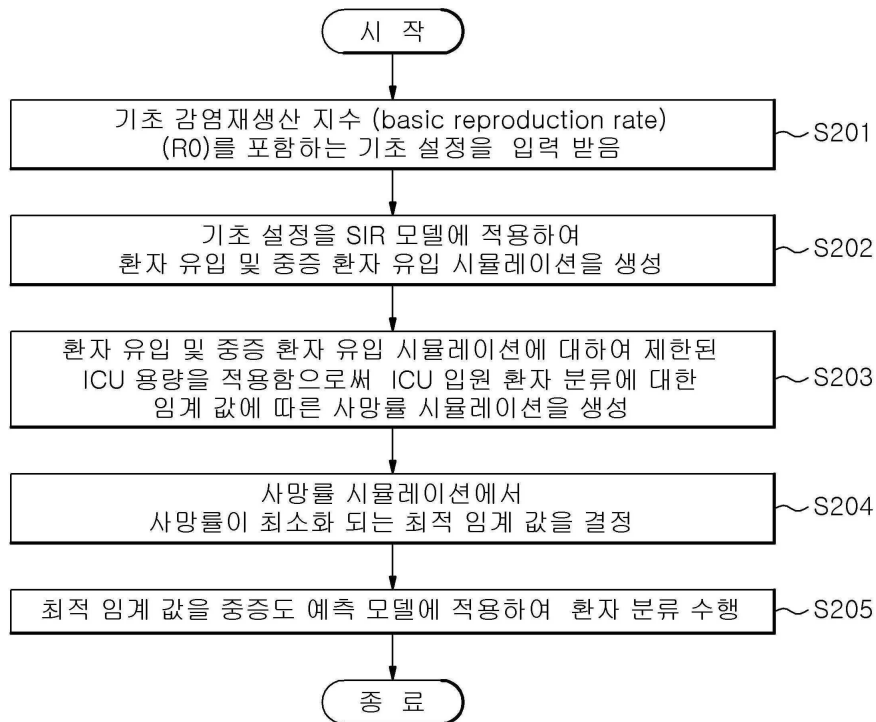
- [0273]
- | | |
|------------|------------|
| 100: 전자 장치 | 110: 입력 장치 |
| 120: 출력 장치 | 130: 메모리 |
| 140: 프로세서 | |

도면

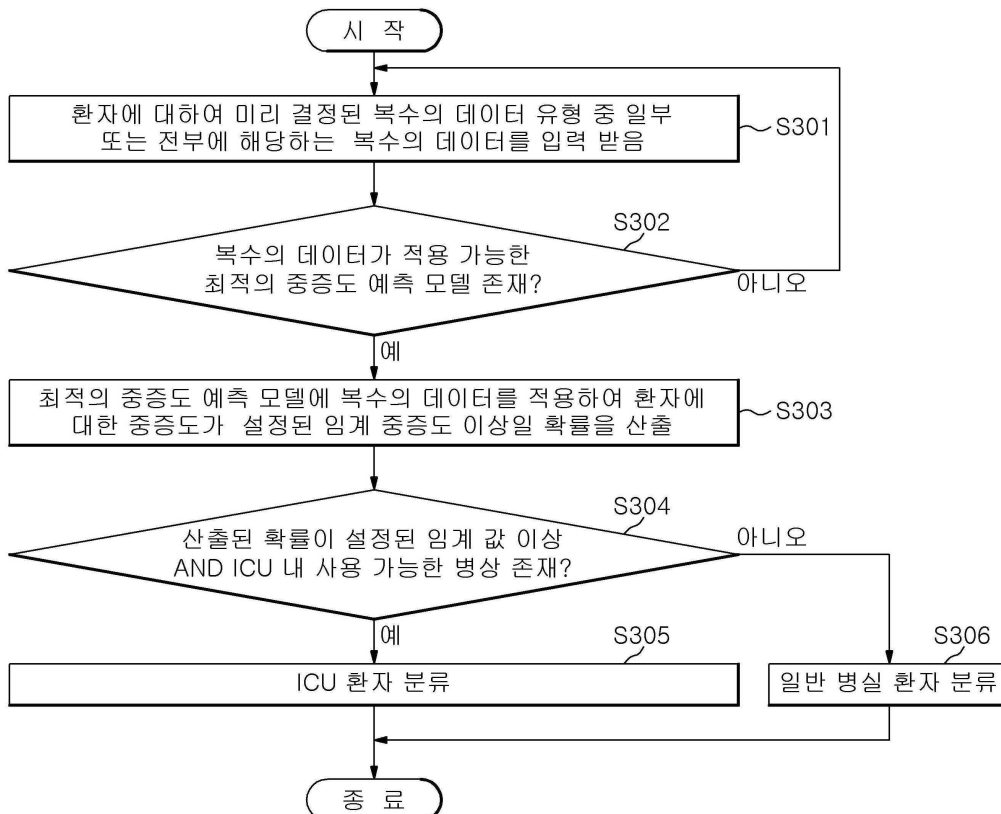
도면1



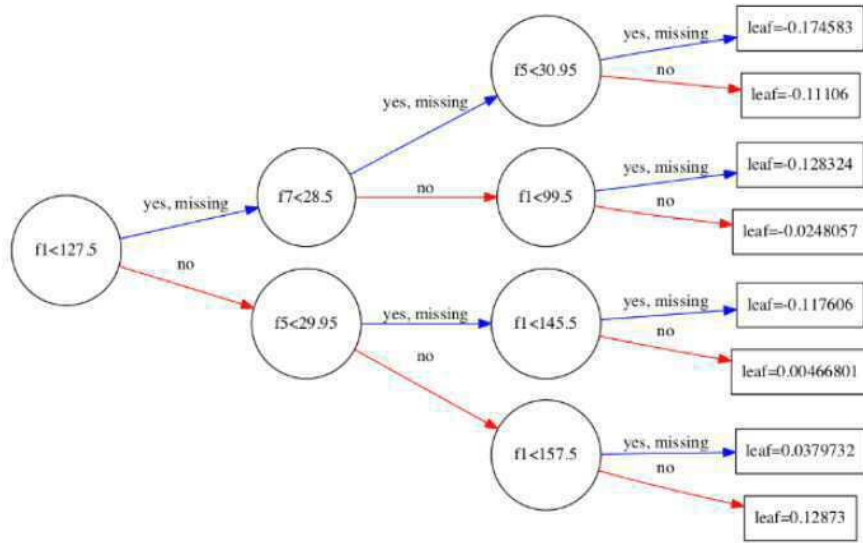
도면2



도면3



도면4



도면5

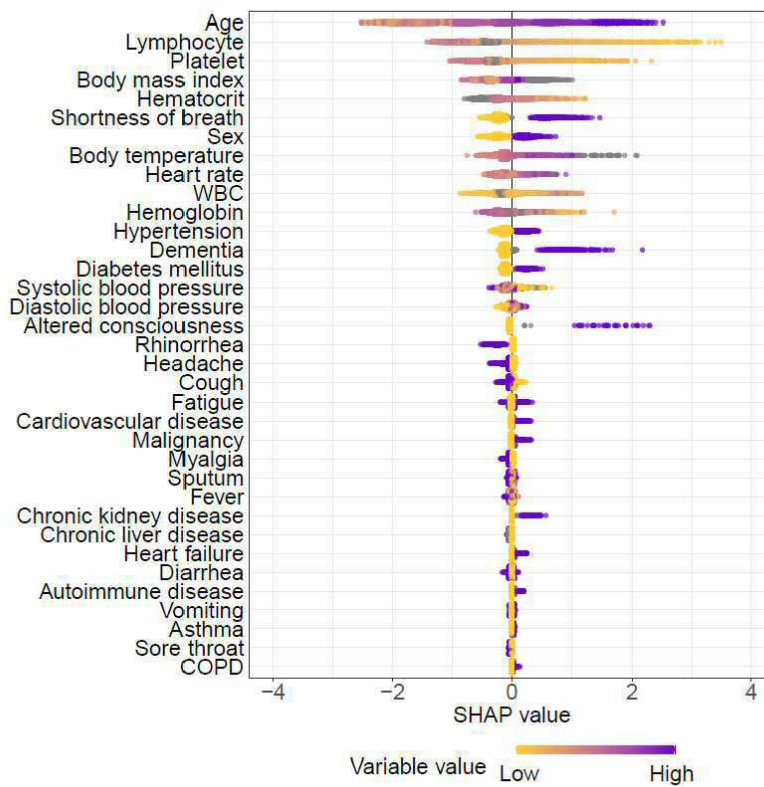
| Variables | Total cohort (n = 5601) | | Train cohort (n = 5037) | | Holdout cohort (n = 564) | | P |
|--------------------------|----------------------------|----------------|----------------------------|----------------|-----------------------------|----------------|-------|
| | M (SD) / n (%) | Missing (%) | M (SD) / n (%) | Missing (%) | M (SD) / n (%) | Missing (%) | |
| Age | | 0.0% | | 0.0% | | 0.0% | 0.407 |
| 0 - 9 | 66 (1.2%) | | 61 (1.2%) | | 5 (0.9%) | | |
| 10 - 19 | 205 (3.7%) | | 185 (3.7%) | | 20 (3.6%) | | |
| 20 - 29 | 1110 (19.8%) | | 988 (19.6%) | | 122 (21.6%) | | |
| 30 - 39 | 564 (10.1%) | | 513 (10.2%) | | 51 (9.0%) | | |
| 40 - 49 | 739 (13.2%) | | 652 (12.9%) | | 87 (15.4%) | | |
| 50 - 59 | 1141 (20.4%) | | 1039 (20.6%) | | 102 (18.1%) | | |
| 60 - 69 | 907 (16.2%) | | 809 (16.1%) | | 98 (17.4%) | | |
| 70 - 79 | 545 (9.7%) | | 495 (9.8%) | | 50 (8.9%) | | |
| ≥ 80 | 324 (5.8%) | | 295 (5.9%) | | 29 (5.1%) | | |
| Sex (male) | 2310 (41.2%) | 0.0% | 2073 (41.2%) | 0.0% | 237 (42.0%) | 0.0% | 0.726 |
| Pregnancy | 19 (0.3%) | 0.4% | 17 (0.3%) | 0.3% | 2 (0.4%) | 0.5% | 1.000 |
| Pregnancy Weeks | 0.05 (1.1) | 0.4% | 0.06 (1.1) | 0.4% | 0.03 (0.5) | 0.5% | 0.399 |
| Body mass index | | 21.4% | | 21.5% | | 20.9% | 0.653 |
| < 18.5 | 259 (4.6%) | | 236 (4.7%) | | 23 (4.1%) | | |
| 18.5 - 22.9 | 1854 (33.1%) | | 1666 (33.1%) | | 188 (33.3%) | | |
| 23.0 - 24.9 | 1035 (18.5%) | | 929 (18.4%) | | 106 (18.8%) | | |
| 25.0 - 29.9 | 1045 (18.7%) | | 938 (18.6%) | | 107 (19.0%) | | |
| ≥ 30 | 207 (3.7%) | | 185 (3.7%) | | 22 (3.9%) | | |
| Systolic blood pressure | | 2.5% | | 2.5% | | 2.7% | 0.600 |
| < 120 | 1306 (23.3%) | | 1177 (23.4%) | | 129 (22.9%) | | |
| 120 - 129 | 1138 (20.3%) | | 1012 (20.1%) | | 126 (22.3%) | | |
| 130 - 139 | 1084 (19.4%) | | 977 (19.4%) | | 107 (19.0%) | | |
| 140 - 159 | 1418 (25.3%) | | 1281 (25.4%) | | 137 (24.3%) | | |
| ≥ 160 | 513 (9.2%) | | 463 (9.2%) | | 50 (8.9%) | | |
| Diastolic blood pressure | | 2.5% | | 2.5% | | 2.7% | 0.036 |
| < 80 | 2102 (37.5%) | | 1878 (37.3%) | | 224 (39.7%) | | |
| 80 - 89 | 1797 (32.1%) | | 1601 (31.8%) | | 196 (34.8%) | | |
| 90 - 99 | 1056 (18.9%) | | 971 (19.3%) | | 85 (15.1%) | | |
| ≥ 100 | 504 (9.0%) | | 460 (9.1%) | | 44 (7.8%) | | |
| Heart rate | 85.82 (15.1) | 2.3% | 85.77 (15.0) | 2.3% | 86.27 (15.4) | 2.5% | 0.474 |
| Body temperature | 36.9 (0.6) | 0.7% | 36.9 (0.6) | 0.8% | 37.0 (0.6) | 0.7% | 0.861 |
| Fever | 1302 (23.3%) | 0.1% | 1168 (23.2%) | 0.1% | 134 (23.8%) | 0.0% | 0.801 |
| Cough | 2331 (41.6%) | 0.1% | 2103 (41.8%) | 0.1% | 228 (40.4%) | 0.0% | 0.575 |
| Sputum | 1611 (28.8%) | 0.1% | 1460 (29.0%) | 0.1% | 151 (26.8%) | 0.0% | 0.293 |
| Sore throat | 872 (15.6%) | 0.1% | 779 (15.5%) | 0.1% | 93 (16.5%) | 0.0% | 0.565 |
| Rhinorrhea | 617 (11.1%) | 0.1% | 560 (11.1%) | 0.1% | 57 (10.1%) | 0.0% | 0.511 |
| Myalgia | 920 (16.4%) | 0.1% | 820 (16.3%) | 0.1% | 100 (17.7%) | 0.0% | 0.411 |
| Fatigue | 233 (4.2%) | 0.1% | 207 (4.1%) | 0.1% | 26 (4.6%) | 0.0% | 0.650 |
| Shortness of breath | 665 (11.9%) | 0.1% | 608 (12.1%) | 0.1% | 57 (10.1%) | 0.0% | 0.194 |
| Headache | 963 (17.2%) | 0.1% | 873 (17.3%) | 0.1% | 90 (16.0%) | 0.0% | 0.446 |
| Altered consciousness | 35 (0.6%) | 0.1% | 31 (0.6%) | 0.1% | 4 (0.7%) | 0.0% | 0.776 |
| Vomiting | 244 (4.4%) | 0.1% | 210 (4.2%) | 0.1% | 34 (6.0%) | 0.0% | 0.052 |
| Diarrhea | 516 (9.2%) | 0.1% | 457 (9.1%) | 0.1% | 59 (10.5%) | 0.0% | 0.315 |
| Diabetes mellitus | 688 (12.3%) | 0.1% | 620 (12.3%) | 0.1% | 68 (12.1%) | 0.0% | 0.916 |
| Hypertension | 1198 (21.4%) | 0.1% | 1087 (21.6%) | 0.1% | 111 (19.7%) | 0.0% | 0.323 |
| Heart failure | 59 (1.1%) | 0.1% | 52 (1.0%) | 0.1% | 7 (1.2%) | 0.0% | 0.808 |
| Cardiovascular disease | 179 (3.2%) | 0.3% | 156 (3.1%) | 0.3% | 23 (4.1%) | 0.4% | 0.259 |
| Asthma | 128 (2.3%) | 0.1% | 118 (2.3%) | 0.1% | 10 (1.8%) | 0.0% | 0.478 |
| COPD | 40 (0.7%) | 0.1% | 38 (0.8%) | 0.1% | 2 (0.4%) | 0.0% | 0.428 |
| Chronic kidney disease | 55 (1.0%) | 0.1% | 48 (1.0%) | 0.1% | 7 (1.2%) | 0.0% | 0.665 |
| Malignancy | 145 (2.6%) | 0.1% | 134 (2.7%) | 0.1% | 11 (2.0%) | 0.0% | 0.386 |
| Chronic liver disease | 83 (1.6%) | 5.8% | 75 (1.6%) | 5.7% | 8 (1.5%) | 6.7% | 1.000 |
| Autoimmune disease | 38 (0.7%) | 5.9% | 32 (0.7%) | 5.8% | 6 (1.1%) | 6.9% | 0.365 |
| Dementia | 224 (4.2%) | 5.9% | 203 (4.3%) | 5.8% | 21 (3.7%) | 6.7% | 0.811 |
| Hemoglobin | 13.28 (1.8) | 27.2% | 13.29 (1.8) | 26.7% | 13.21 (1.8) | 31.6% | 0.412 |
| Hematocrit | 39.24 (5.0) | 27.2% | 39.26 (4.9) | 26.7% | 39.09 (5.2) | 31.7% | 0.558 |
| Lymphocyte | 29.15 (11.7) | 27.6% | 29.25 (11.7) | 27.1% | 28.23 (11.0) | 32.1% | 0.088 |
| Platelet | 23697 (82897) | 27.1% | 236776 (82534) | 26.7% | 235943 (86395) | 31.4% | 0.856 |
| WBC | 6126 (2824) | 27.1% | 6121 (2841) | 26.7% | 6167 (2666) | 31.4% | 0.748 |
| CSS ≥6 | 271 (4.8%) | 0.0% | 242 (4.8%) | 0.0% | 29 (5.1%) | 0.0% | 0.802 |

도면6

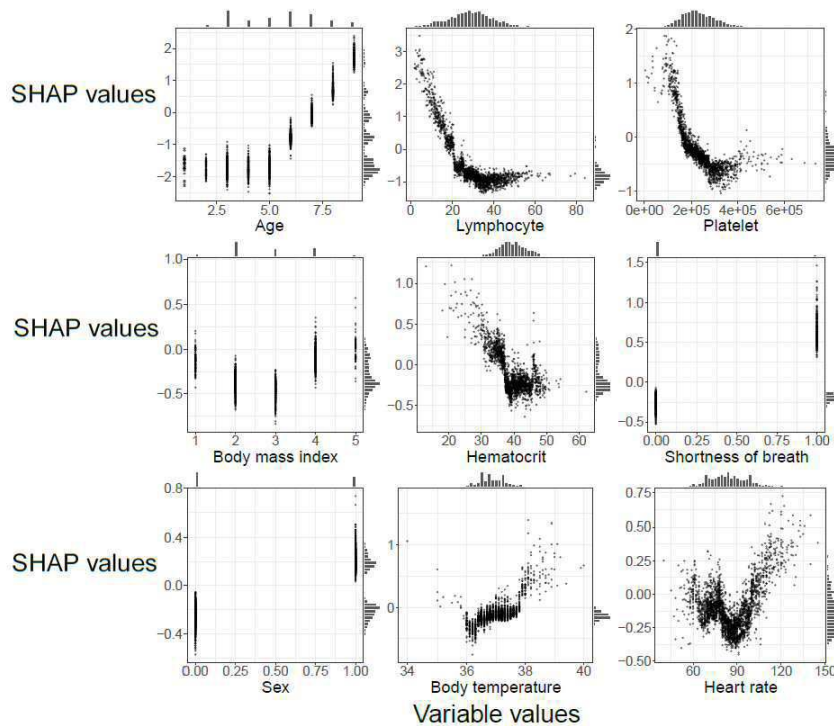
| CSS | Model | AUROC | Specificity | Sensitivity | Accuracy | PPV | NPV | P |
|-----|-------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|-------|
| 5 | LR | 0.915 (0.882 - 0.941) | 0.901 (0.891 - 0.908) | 0.766 (0.695 - 0.830) | 0.893 (0.882 - 0.905) | 0.300 (0.270 - 0.330) | 0.986 (0.982 - 0.990) | 0.030 |
| | XGB | 0.950 (0.936 - 0.961) | 0.900 (0.891 - 0.908) | 0.882 (0.838 - 0.919) | 0.899 (0.890 - 0.907) | 0.333 (0.312 - 0.356) | 0.993 (0.990 - 0.995) | |
| 6 | LR | 0.938 (0.911 - 0.959) | 0.900 (0.888 - 0.912) | 0.814 (0.737 - 0.881) | 0.896 (0.885 - 0.907) | 0.273 (0.246 - 0.302) | 0.991 (0.987 - 0.994) | 0.039 |
| | XGB | 0.965 (0.958 - 0.972) | 0.900 (0.892 - 0.908) | 0.905 (0.868 - 0.942) | 0.901 (0.892 - 0.909) | 0.314 (0.295 - 0.335) | 0.995 (0.993 - 0.997) | |
| 7 | LR | 0.938 (0.903 - 0.963) | 0.900 (0.888 - 0.912) | 0.857 (0.790 - 0.924) | 0.898 (0.887 - 0.910) | 0.259 (0.234 - 0.289) | 0.994 (0.993 - 0.997) | 0.060 |
| | XGB | 0.966 (0.959 - 0.973) | 0.900 (0.892 - 0.909) | 0.902 (0.862 - 0.938) | 0.900 (0.892 - 0.909) | 0.297 (0.278 - 0.318) | 0.995 (0.993 - 0.997) | |
| 8 | LR | 0.940 (0.907 - 0.964) | 0.900 (0.889 - 0.911) | 0.867 (0.796 - 0.929) | 0.899 (0.887 - 0.909) | 0.248 (0.223 - 0.273) | 0.994 (0.991 - 0.997) | 0.062 |
| | XGB | 0.968 (0.961 - 0.975) | 0.900 (0.892 - 0.908) | 0.903 (0.861 - 0.940) | 0.900 (0.892 - 0.908) | 0.288 (0.270 - 0.307) | 0.995 (0.993 - 0.997) | |

Data are value (95% confidence interval)

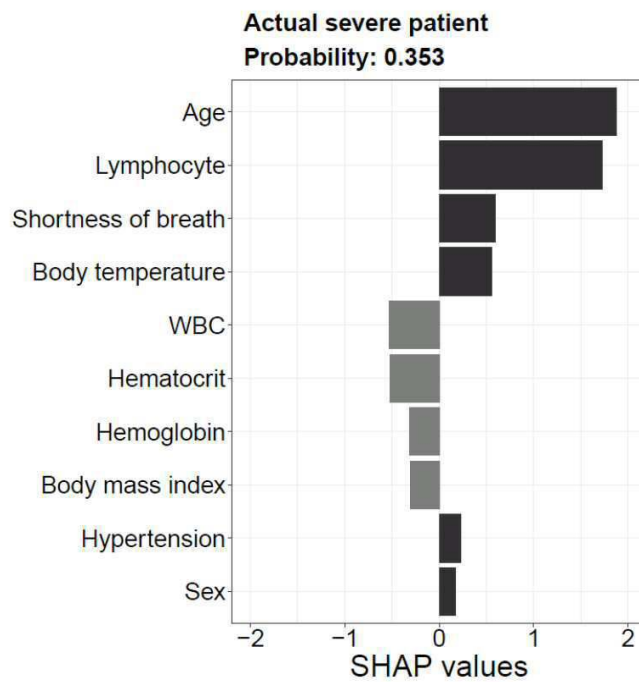
도면7



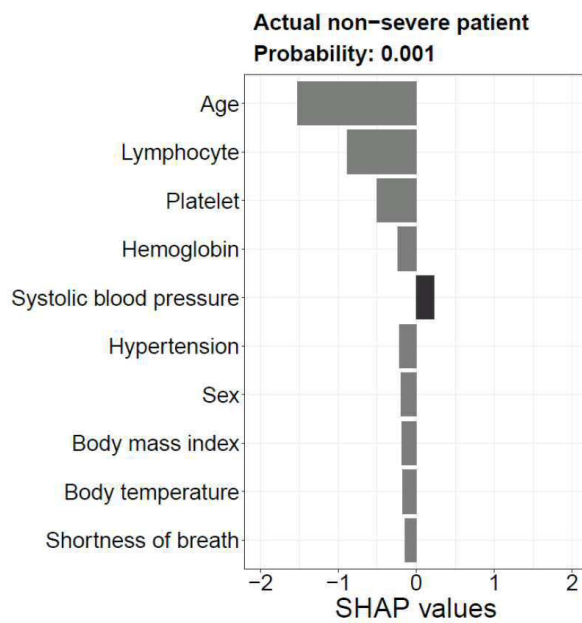
도면8



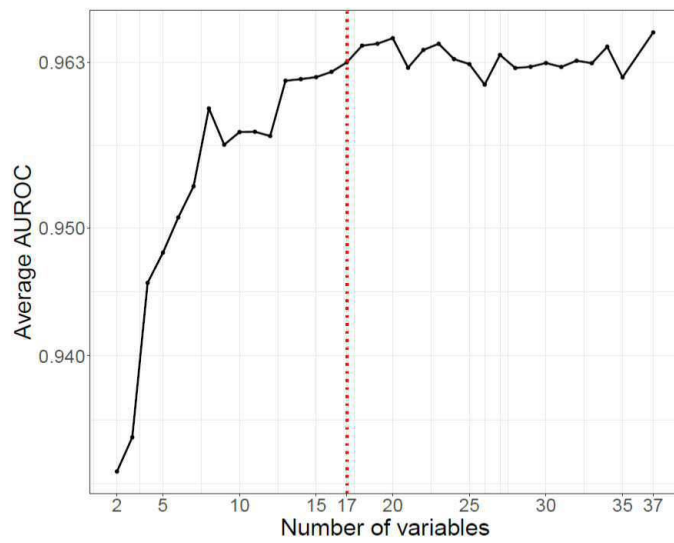
도면9



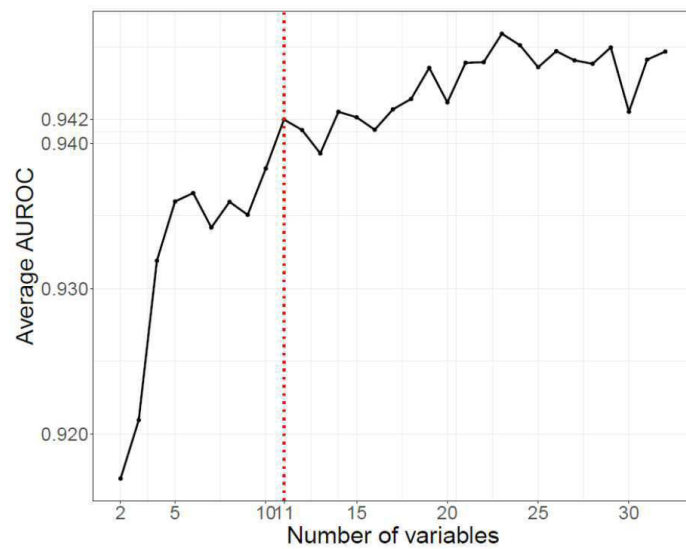
도면10



도면11



도면12



도면13

| Number of variables | F+L+ model | | F+L- model | |
|---------------------|------------|-------|------------|-------|
| | AUROC | P | AUROC | P |
| 1 | 0.889 | 0.000 | 0.889 | 0.000 |
| 2 | 0.931 | 0.000 | 0.917 | 0.000 |
| 3 | 0.934 | 0.000 | 0.921 | 0.000 |
| 4 | 0.946 | 0.000 | 0.932 | 0.001 |
| 5 | 0.948 | 0.000 | 0.936 | 0.008 |
| 6 | 0.951 | 0.000 | 0.937 | 0.009 |
| 7 | 0.953 | 0.000 | 0.934 | 0.001 |
| 8 | 0.959 | 0.003 | 0.936 | 0.007 |
| 9 | 0.956 | 0.000 | 0.935 | 0.000 |
| 10 | 0.957 | 0.000 | 0.938 | 0.007 |
| 11 | 0.958 | 0.000 | 0.942 | 0.137 |
| 12 | 0.957 | 0.000 | 0.941 | 0.043 |
| 13 | 0.961 | 0.018 | 0.939 | 0.014 |
| 14 | 0.962 | 0.018 | 0.942 | 0.087 |
| 15 | 0.962 | 0.015 | 0.942 | 0.063 |
| 16 | 0.962 | 0.021 | 0.941 | 0.015 |
| 17 | 0.963 | 0.080 | 0.942 | 0.108 |
| 18 | 0.964 | 0.252 | 0.943 | 0.051 |
| 19 | 0.964 | 0.379 | 0.945 | 0.435 |
| 20 | 0.965 | 0.392 | 0.943 | 0.024 |
| 21 | 0.963 | 0.016 | 0.946 | 0.556 |
| 22 | 0.964 | 0.352 | 0.946 | 0.640 |
| 23 | 0.964 | 0.510 | 0.948 | 0.551 |
| 24 | 0.963 | 0.107 | 0.947 | 0.988 |
| 25 | 0.963 | 0.077 | 0.945 | 0.517 |
| 26 | 0.961 | 0.002 | 0.946 | 0.693 |
| 27 | 0.964 | 0.141 | 0.946 | 0.305 |
| 28 | 0.962 | 0.014 | 0.945 | 0.437 |
| 29 | 0.963 | 0.027 | 0.947 | 0.672 |
| 30 | 0.963 | 0.043 | 0.942 | 0.004 |
| 31 | 0.963 | 0.116 | 0.946 | 0.686 |
| 32 | 0.963 | 0.054 | 0.946 | - |
| 33 | 0.963 | 0.173 | - | - |
| 34 | 0.964 | 0.378 | - | - |
| 35 | 0.962 | 0.009 | - | - |
| 37 | 0.965 | - | - | - |

도면14

| Model | Number of variables | AUROC | Specificity | Sensitivity | Accuracy | PPV | NPV |
|-------|---------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| F+L+ | 37 | 0.965 (0.958-0.972) | 0.900 (0.892-0.909) | 0.905 (0.868-0.942) | 0.900 (0.892-0.908) | 0.314 (0.295-0.335) | 0.995 (0.993-0.997) |
| F-L+ | 17 | 0.963 (0.955-0.971) | 0.900 (0.892-0.908) | 0.884 (0.839-0.921) | 0.899 (0.891-0.907) | 0.309 (0.289-0.329) | 0.994 (0.991-0.996) |
| F+L- | 32 | 0.946 (0.936-0.956) | 0.900 (0.891-0.908) | 0.839 (0.793-0.884) | 0.897 (0.888-0.905) | 0.297 (0.276-0.319) | 0.991 (0.988-0.994) |
| F-L- | 11 | 0.942 (0.931-0.953) | 0.901 (0.892-0.909) | 0.810 (0.756-0.860) | 0.896 (0.888-0.904) | 0.291 (0.270-0.313) | 0.989 (0.987-0.992) |

도면15

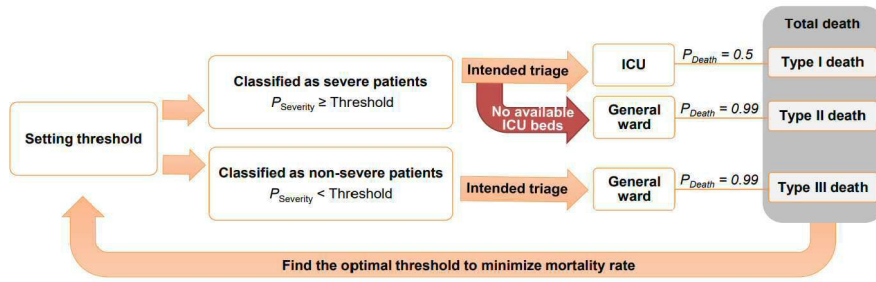
| Model | Number of variables | AUROC | Specificity | Sensitivity | Accuracy | PPV | NPV | P |
|-------|---------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|-------|
| F+L+ | 37 | 0.958 (0.924-0.991) | 0.882 (0.854-0.908) | 0.966 (0.897-1.000) | 0.887 (0.860-0.911) | 0.309 (0.262-0.367) | 0.998 (0.994-1.000) | 0.663 |
| F-L+ | 17 | 0.949 (0.906-0.99) | 0.888 (0.862-0.914) | 0.966 (0.897-1.000) | 0.892 (0.867-0.917) | 0.319 (0.270-0.378) | 0.998 (0.994-1.000) | 0.541 |
| F+L- | 32 | 0.943 (0.901-0.985) | 0.907 (0.882-0.931) | 0.828 (0.690-0.966) | 0.902 (0.878-0.926) | 0.325 (0.265-0.400) | 0.990 (0.982-0.998) | 0.889 |
| F-L- | 11 | 0.941 (0.903-0.978) | 0.912 (0.890-0.935) | 0.828 (0.690-0.931) | 0.908 (0.885-0.931) | 0.338 (0.274-0.415) | 0.990 (0.982-0.996) | 0.950 |

Data are value (95% confidence interval)

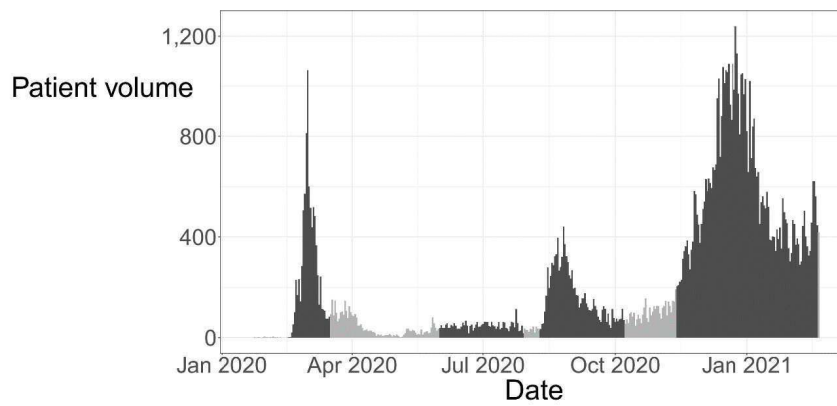
도면16

| Order | With laboratory variable | Without laboratory variable |
|-------|---------------------------------------|---------------------------------------|
| 1 | Age | Age |
| 2 | Lymphocyte | Shortness of breath |
| 3 | Platelet | Body mass index |
| 4 | Body mass index | Sex |
| 5 | Hematocrit | Body temperature |
| 6 | Shortness of breath | Heart rate |
| 7 | Sex | Hypertension |
| 8 | Body temperature | Systolic blood pressure |
| 9 | Heart rate | Diabetes mellitus |
| 10 | White blood cell | Dementia |
| 11 | Hemoglobin | Diastolic blood pressure |
| 12 | Hypertension | Rhinorrhea |
| 13 | Dementia | Fever |
| 14 | Diabetes mellitus | Cough |
| 15 | Systolic blood pressure | Headache |
| 16 | Diastolic blood pressure | Sore throat |
| 17 | Altered consciousness | Altered consciousness |
| 18 | Rhinorrhea | Malignancy |
| 19 | Headache | Sputum |
| 20 | Cough | Myalgia |
| 21 | Fatigue | Cardiovascular disease |
| 22 | Cardiovascular disease | Chronic kidney disease |
| 23 | Malignancy | Diarrhea |
| 24 | Myalgia | Vomiting |
| 25 | Sputum | Chronic liver disease |
| 26 | Fever | Fatigue |
| 27 | Chronic kidney disease | Asthma |
| 28 | Chronic liver disease | Heart failure |
| 29 | Heart failure | Chronic obstructive pulmonary disease |
| 30 | Diarrhea | Autoimmune disease |
| 31 | Autoimmune disease | Pregnancy |
| 32 | Vomiting | Pregnancy Weeks |
| 33 | Asthma | N/A |
| 34 | Sore throat | N/A |
| 35 | Chronic obstructive pulmonary disease | N/A |
| 37 | Pregnancy | N/A |
| 37 | Pregnancy Weeks | N/A |

도면17

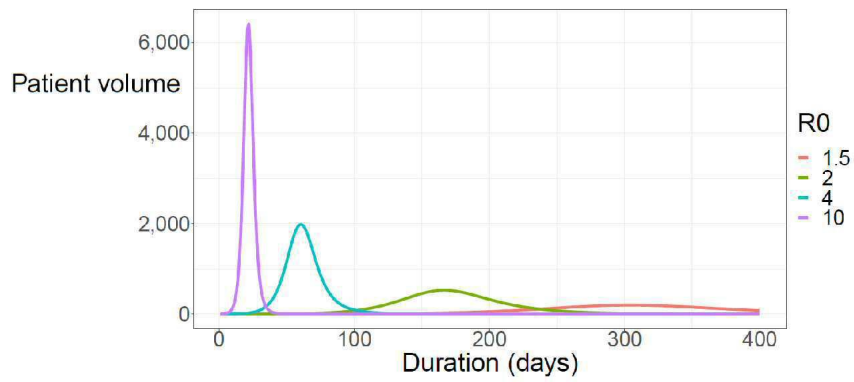


도면18



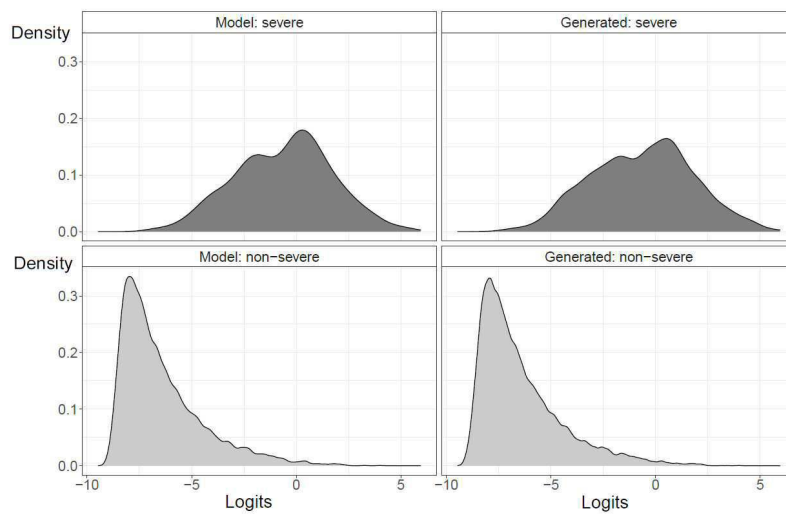
| | H 1 | H 2 | H 3 | H 4 |
|--|---|--|--|---|
| Period | 17 th Feb - 17 th Mar | 1 st Jun - 30 th Jul | 10 th Aug - 8 th Oct | 13 th Nov - 20 th Feb |
| Duration (days) | 30 | 60 | 60 | 100 |
| Cumulative patients | 8,384 | 2,802 | 9,850 | 58,859 |
| Maximum daily patients (= maximum height) | 1,062 | 113 | 441 | 1,240 |

도면19

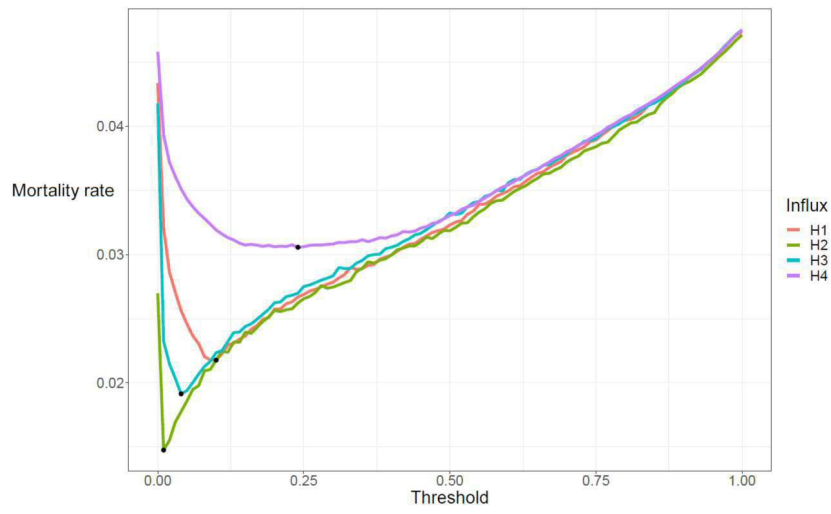


| R0 | 1.5 | 2 | 4 | 10 |
|--|--------|--------|--------|--------|
| Duration (days) | 677 | 416 | 198 | 65 |
| Cumulative patients | 34,965 | 47,802 | 58,799 | 59,993 |
| Maximum daily patients (= maximum height) | 198 | 526 | 1,980 | 6,407 |

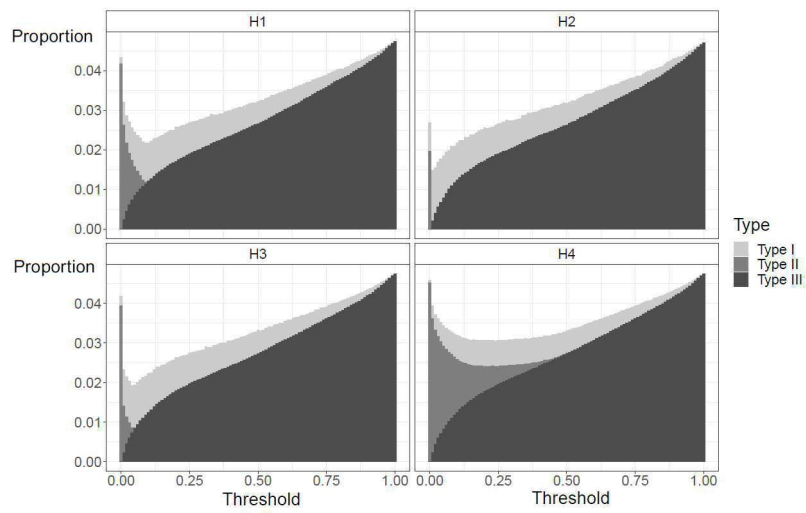
도면20



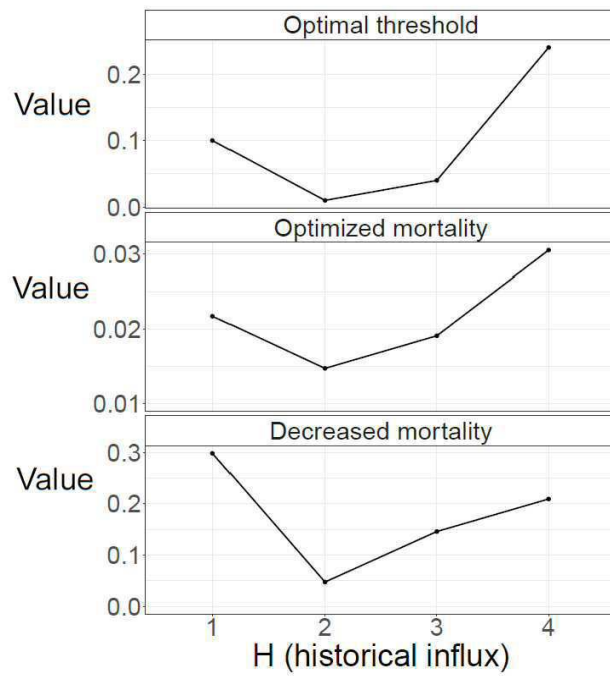
도면21



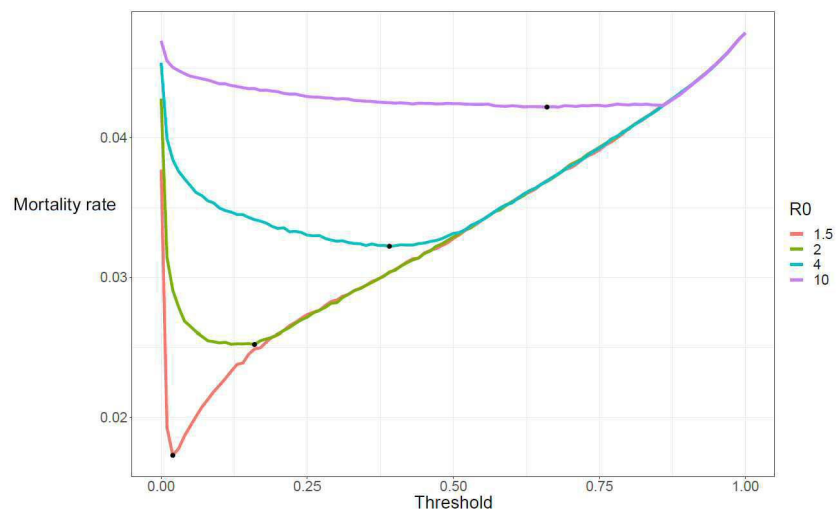
도면22



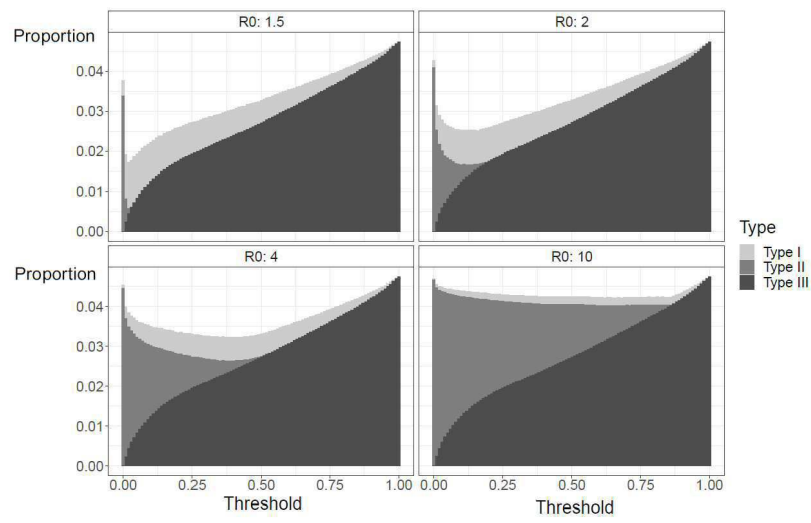
도면23



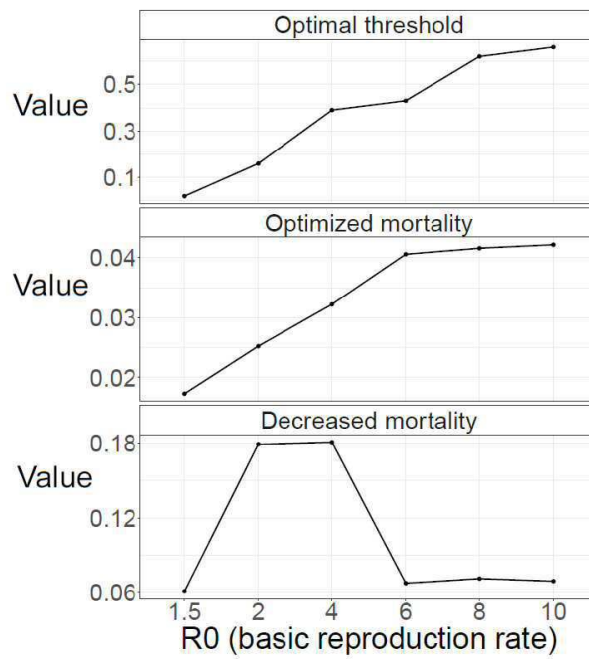
도면24



도면25



도면26



도면27

| Influx | Optimal threshold | Optimized mortality | Decreased mortality |
|----------|-------------------|---------------------|---------------------|
| H1 | 0.1 | 0.022 | 0.298 |
| H2 | 0.01 | 0.015 | 0.047 |
| H3 | 0.04 | 0.019 | 0.146 |
| H4 | 0.24 | 0.031 | 0.209 |
| R0 = 1.5 | 0.02 | 0.017 | 0.061 |
| R0 = 2 | 0.16 | 0.025 | 0.179 |
| R0 = 4 | 0.39 | 0.032 | 0.181 |
| R0 = 6 | 0.43 | 0.041 | 0.068 |
| R0 = 8 | 0.62 | 0.042 | 0.071 |
| R0 = 10 | 0.66 | 0.042 | 0.069 |