



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0119460
(43) 공개일자 2023년08월16일

(51) 국제특허분류(Int. Cl.)

G16H 50/30 (2018.01) G06N 20/00 (2019.01)
G16B 40/00 (2019.01) G16H 10/60 (2018.01)
G16H 50/20 (2018.01)

(52) CPC특허분류

G16H 50/30 (2018.01)
G06N 20/00 (2021.08)

(21) 출원번호 10-2022-0015635

(22) 출원일자 2022년02월07일

심사청구일자 2022년02월07일

(71) 출원인

재단법인 아산사회복지재단

서울특별시 송파구 올림픽로43길 88 (풍납동)

울산대학교 산학협력단

울산광역시 남구 대학로 93(무거동)

(72) 발명자

전태준

서울특별시 송파구 풍성로6길 15, 101동 708호(풍납동, 신동아아파트)

최은지

서울특별시 강남구 연주로 117, 9동 408호(도곡동, 우성4차아파트)

(뒷면에 계속)

(74) 대리인

특허법인 무한

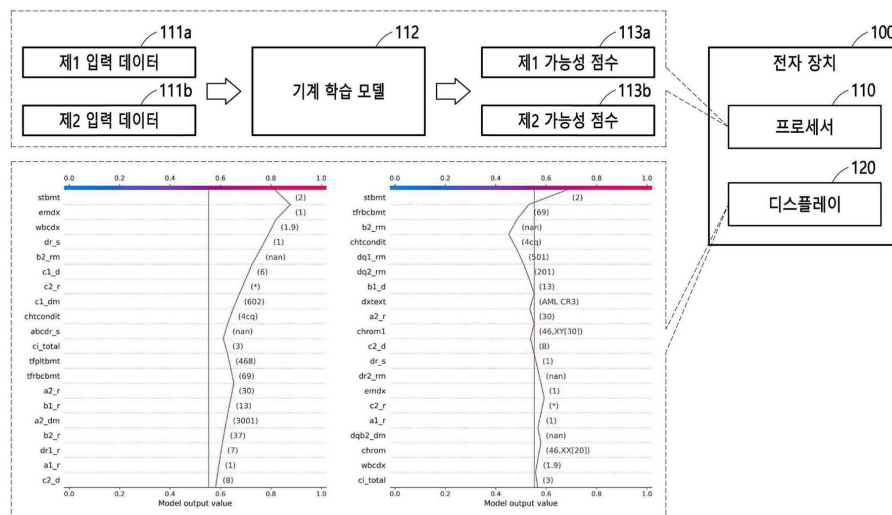
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 기계 학습 모델을 이용하여 동종 조혈 세포 이식 이후 수여자의 사망을 예측하기 위한 방법 및 장치

(57) 요약

일 실시예에 따른 전자 장치는, 공여자(donor)에 관한 제1 의료 피처들(medical features), 수여자(recipient)에 관한 제2 의료 피처들, 및 상기 공여자와 상기 수여자 간의 관계(relationship)에 관한 제3 의료 피처들을 가지는 제1 입력 데이터에 기계 학습 모델을 적용하여 동종 조혈 세포 이식(allogeneic stem cell transplantation)의 수여자가 미리 정의된 기간 내 사망할 제1 가능성 점수를 획득하고, 상기 제1 입력 데이터의 의료 피처들 중 적어도 일부의 값을 변경함으로써 획득된 제2 입력 데이터에 상기 기계 학습 모델을 적용하여 제2 가능성 점수 획득하는 프로세서 및 상기 제1 가능성 점수 및 상기 제2 가능성 점수를 출력하는 디스플레이를 포함할 수 있다.

대표도



(52) CPC특허분류

G16B 40/00 (2019.02)

G16H 10/60 (2021.08)

G16H 50/20 (2018.01)

(72) 발명자

이제환

서울특별시 노원구 중계로 233, 101동 305호(중계동, 청구3차아파트)

김영학

서울특별시 강남구 남부순환로 2914, 3동 1102호(대치동, 개포1차우성아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호 1465034378

과제번호 HR21C0198010021

부처명 보건복지부

과제관리(전문)기관명 한국보건산업진흥원

연구사업명 연구중심병원육성(R&D)

연구과제명 의료현장 통합 지식베이스 및 인공지능형 가상 임상시험 시스템 구축

기 여 율 1/1

과제수행기관명 서울아산병원

연구기간 2021.07.01 ~ 2021.12.31

명세서

청구범위

청구항 1

프로세서에 의하여 수행되는 방법에 있어서,

공여자(donor)에 관한 제1 의료 피처들(medical features), 수여자(recipient)에 관한 제2 의료 피처들, 및 상기 공여자와 상기 수여자 간의 관계(relationship)에 관한 제3 의료 피처들을 가지는 제1 입력 데이터에 기계 학습 모델을 적용하여 동종 조혈 세포 이식(allogeneic stem cell transplantation)의 수여자가 미리 정의된 기간 내 사망할 제1 가능성 점수를 획득하는 단계;

상기 제1 입력 데이터의 의료 피처들 중 적어도 일부의 값을 변경함으로써 획득된 제2 입력 데이터에 상기 기계 학습 모델을 적용하여 제2 가능성 점수 획득하는 단계; 및

상기 제1 가능성 점수 및 상기 제2 가능성 점수를 출력하는 단계를 포함하는 방법.

청구항 2

제1항에 있어서,

상기 제1 입력 데이터 및 상기 제2 입력 데이터의 의료 피처들의 값들 각각의 상기 제1 가능성 점수 및 제2 가능성 점수에 대한 영향력을 나타내는 피처 기여도(feature contribution)를 계산하는 단계; 및

상기 계산된 피처 기여도를 출력하는 단계를 더 포함하는 방법.

청구항 3

제2항에 있어서,

상기 피처 기여도를 출력하는 단계는,

상기 피처 기여도의 크기에 따라 정렬된 상기 제1 입력 데이터의 상기 의료 피처들의 상기 피처 기여도를 출력하는 단계; 및

상기 피처 기여도의 크기에 따라 정렬된 상기 제2 입력 데이터의 상기 의료 피처들의 상기 피처 기여도를 출력하는 단계를 포함하는, 방법.

청구항 4

제1항에 있어서,

상기 제2 가능성 점수 획득하는 단계는,

다른 공여자에 대한 의료 피처 값들을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 상기 제1 의료 피처들 및 제3 의료 피처들 중 적어도 일부의 값을 상기 공여자에 대응하는 값에서 상기 다른 공여자에 대응하는 값으로 변경함으로써 상기 제2 입력 데이터를 획득하는 단계; 및

상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득하는 단계를 포함하

는,
방법.

청구항 5

제1항에 있어서,

상기 제2 가능성 점수 획득하는 단계는,

다른 수여자에 대한 의료 피쳐 값들을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 상기 제2 의료 피쳐들 및 상기 제3 의료 피쳐들 중 적어도 일부의 값을 상기 수여자에 대응하는 값에서 상기 다른 수여자에 대응하는 값으로 변경함으로써 상기 제2 입력 데이터를 획득하는 단계; 및

상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득하는 단계를 포함하는,

방법.

청구항 6

제1항에 있어서,

상기 제2 가능성 점수 획득하는 단계는,

동종 조혈 세포 이식의 계획(planning)에 따라 조정된 의료 피쳐 값을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 조정가능한 의료 피쳐들 중 적어도 일부의 값을 상기 조정된 의료 피쳐 값으로 변경함으로써 상기 제2 입력 데이터를 획득하는 단계; 및

상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득하는 단계를 포함하는,

방법.

청구항 7

제1항에 있어서,

복수의 의료 피쳐들을 포함하는 임시 입력 데이터에 기초하여 트레이닝된 임시 기계 학습 모델을 이용하여 상기 복수의 의료 피쳐들 중 상기 기계 학습 모델의 입력 데이터의 의료 피쳐들을 결정하는 단계; 및

상기 입력 데이터의 의료 피쳐들을 가지는 트레이닝 입력 데이터에 기초하여 트레이닝된 상기 기계 학습 모델을 획득하는 단계

를 더 포함하는 방법.

청구항 8

제7항에 있어서,

상기 기계 학습 모델의 입력 데이터의 의료 피쳐들을 결정하는 단계는,

상기 임시 입력 데이터에 기초하여 상기 임시 기계 학습 모델을 트레이닝시키는 단계;

상기 트레이닝된 임시 기계 학습 모델에 대한 상기 임시 입력 데이터의 복수의 의료 피쳐들의 피쳐 중요도(feature importance)에 기초하여 상기 임시 입력 데이터의 상기 복수의 의료 피쳐들 중 일부를 상기 임시 입력 데이터에서 제거하는 단계; 및

상기 임시 기계 학습 모델의 트레이닝 및 상기 임시 입력 데이터에 대한 의료 피쳐 제거를 반복하는 단계를 포함하는,
방법.

청구항 9

제8항에 있어서,

상기 임시 기계 학습 모델의 트레이닝 및 상기 임시 입력 데이터의 의료 피쳐 제거를 반복하는 단계는,

상기 임시 기계 학습 모델의 트레이닝 및 상기 임시 입력 데이터의 의료 피쳐 제거와 함께 상기 임시 기계 학습 모델의 정확도 계산을 반복하는 단계; 및

상기 계산된 정확도에 기초하여 상기 임시 입력 데이터의 의료 피쳐들을 상기 기계 학습 모델의 상기 입력 데이터의 상기 의료 피쳐들로 결정하는 단계를 포함하는,

방법.

청구항 10

하드웨어와 결합되어 제1항 내지 제9항 중 어느 하나의 항의 방법을 실행시키기 위하여 컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램.

청구항 11

공여자(donor)에 관한 제1 의료 피쳐들(medical features), 수여자(recipient)에 관한 제2 의료 피쳐들, 및 상기 공여자와 상기 수여자 간의 관계(relationship)에 관한 제3 의료 피쳐들을 가지는 제1 입력 데이터에 기계 학습 모델을 적용하여 동종 조혈 세포 이식(allogeneic stem cell transplantation)의 수여자가 미리 정의된 기간 내 사망할 제1 가능성 점수를 획득하고, 상기 제1 입력 데이터의 의료 피쳐들 중 적어도 일부의 값을 변경함으로써 획득된 제2 입력 데이터에 상기 기계 학습 모델을 적용하여 제2 가능성 점수 획득하는 프로세서; 및

상기 제1 가능성 점수 및 상기 제2 가능성 점수를 출력하는 디스플레이

를 포함하는 전자 장치.

청구항 12

제11항에 있어서,

상기 프로세서는,

상기 제1 입력 데이터 및 상기 제2 입력 데이터의 의료 피쳐들의 값들 각각의 상기 제1 가능성 점수 및 제2 가능성 점수에 대한 영향력을 나타내는 피쳐 기여도(feature contribution)를 계산하고,

상기 디스플레이는,

상기 계산된 피쳐 기여도를 출력하는,

전자 장치.

청구항 13

제12항에 있어서,

상기 프로세서는,

상기 피처 기여도의 크기에 따라 정렬된 상기 제1 입력 데이터의 상기 의료 피처들의 상기 피처 기여도를 출력하고,

상기 피처 기여도의 크기에 따라 정렬된 상기 제2 입력 데이터의 상기 의료 피처들의 상기 피처 기여도를 출력하는,

전자 장치.

청구항 14

제11항에 있어서,

상기 프로세서는,

다른 공여자에 대한 의료 피처 값들을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 상기 제1 의료 피처들 및 제3 의료 피처들 중 적어도 일부의 값을 상기 공여자에 대응하는 값에서 상기 다른 공여자에 대응하는 값으로 변경함으로써 상기 제2 입력 데이터를 획득하고,

상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득하는,

전자 장치.

청구항 15

제11항에 있어서,

상기 프로세서는,

다른 수여자에 대한 의료 피처 값들을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 상기 제2 의료 피처들 및 상기 제3 의료 피처들 중 적어도 일부의 값을 상기 수여자에 대응하는 값에서 상기 다른 수여자에 대응하는 값으로 변경함으로써 상기 제2 입력 데이터를 획득하고,

상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득하는,

전자 장치.

청구항 16

제11항에 있어서,

상기 프로세서는,

동종 조혈 세포 이식의 계획(planning)에 따라 조정된 의료 피처 값을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 조정가능한 의료 피처들 중 적어도 일부의 값을 상기 조정된 의료 피처 값으로 변경함으로써 상기 제2 입력 데이터를 획득하고,

상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득하는,

전자 장치.

청구항 17

제11항에 있어서,

상기 프로세서는,

복수의 의료 피쳐들을 포함하는 임시 입력 데이터에 기초하여 트레이닝된 임시 기계 학습 모델을 이용하여 상기 복수의 의료 피쳐들 중 상기 기계 학습 모델의 입력 데이터의 의료 피쳐들을 결정하고,

상기 입력 데이터의 의료 피쳐들을 가지는 트레이닝 입력 데이터에 기초하여 트레이닝된 상기 기계 학습 모델을 획득하는,

전자 장치.

청구항 18

제17항에 있어서,

상기 프로세서는,

상기 임시 입력 데이터에 기초하여 상기 임시 기계 학습 모델을 트레이닝시키고,

상기 트레이닝된 임시 기계 학습 모델에 대한 상기 임시 입력 데이터의 복수의 의료 피쳐들의 피쳐 중요도(feature importance)에 기초하여 상기 임시 입력 데이터의 상기 복수의 의료 피쳐들 중 일부를 상기 임시 입력 데이터에서 제거하며,

상기 임시 기계 학습 모델의 트레이닝 및 상기 임시 입력 데이터에 대한 의료 피쳐 제거를 반복하는,

전자 장치.

청구항 19

제18항에 있어서,

상기 프로세서는,

상기 임시 기계 학습 모델의 트레이닝 및 상기 임시 입력 데이터의 의료 피쳐 제거와 함께 상기 임시 기계 학습 모델의 정확도 계산을 반복하고,

상기 계산된 정확도에 기초하여 상기 임시 입력 데이터의 의료 피쳐들을 상기 기계 학습 모델의 상기 입력 데이터의 상기 의료 피쳐들로 결정하는,

전자 장치.

발명의 설명

기술 분야

[0001] 이하, 동종 조혈 세포 이식 이후 생존 및 사망을 예측하기 위한 기계 학습 모델에 관한 기술이 개시된다.

배경 기술

[0002] 동종 조혈 세포 이식(allogeneic hematopoietic cell transplantation; allogeneic HCT) 후 생존을 예측하기 위해 개발된 스코어링 시스템들(Scoring systems)은 차선의(suboptimal) 예측력을 나타낼 수 있다. 이식 후의 결과들에 영향을 미치는 다양한 팩터들이 존재할 수 있다.

[0003] 동종 조혈 세포 이식(HCT)은 현재 널리 사용되는 악성 혈액 종양 환자를 위한 잠재적인 치료 옵션일 수 있다. 동종 HCT의 사용 증가는 개선된 대체 공여자 가용성(alternative donor availability), 감소된 강도 조절 요법(reduced-intensity conditioning regimens), 이식 관련 독성 예방의 발전, 및 일반적인 지지 치료를 포함한 다수의 팩터들에 기인할 수 있다. 이러한 발전에도 불구하고, 동종 HCT는 여전히 상당한 비율의 합병증(complications), 치료 관련 사망률(treatment-related mortality) 및 재발(relapse)과 관련될 수 있다.

[0004] 이식 적격성(transplant eligibility)에 관한 결정을 내리기 전에 이식 결과를 보다 더 정확하게 예측하기 위해

여, 동종 HCT 후 생존(survival)에 대한 여러 예후 스코어링 시스템들(prognostic scoring systems)은 개발되었을 수 있다. 스코어는 HCT-Comorbidity Index(HCT-CI), EBMT(European Group for Blood and Marrow Transplantation) 스코어, PAM(Pretransplantation Assessment of Mortality) 스코어를 포함할 수 있다. 대부분의 예후 스코어링 시스템은 HCT 수여자들의 생존을 위한 콕스 비례 위험 모델(Cox proportional hazard models)과 같은 모수적(parametric) 통계 방법론들을 사용하여 개발되었을 수 있다. 스코어링 시스템의 변수(variable)는 주로 연령, 동반 질환들(comorbidities), 수행 상태(performance status), 진단에서 HCT까지의 시간 및 질병 상태와 같은 수여자 팩터를 포함할 수 있다. EBMT 스코어에서, 공여자 유형(예: 인간 백혈구 항원(human leukocyte antigen; HLA) 일치 형제, 및 HLA 일치 타인), 성별 일치, 및 거대 시토메갈로 혈청 상태(cytomegalovirus serostatus)를 포함한 여러가지 공여자 팩터들은 고려될 수 있다. 그러나 예측 모델은, 보고된 정확도는 수신기 작동 특성 곡선 아래 영역(AUC)의 범위가 0.6 및 0.7 사이라는 점에서 차선(suboptimal)일 수 있다.

[0005] 최근, 딥 러닝 기반 예측 모델들을 이용하여 조기 사망률(early mortality), 이식편대숙주병(graft-versus-host disease; GVHD), 또는 재발과 관련하여 다양한 임상 환경들에서 이식 관련 결과들을 보다 정확하게 예측하려는 시도가 이루어질 수 있다.

[0006] 급성 백혈병(acute leukemia; AL)-혈액 및 골수 이식에 대한 유럽 그룹(European Group for Blood and Marrow Transplantation; EBMT) 스코어는 동종 HCT 후 100일 사망률(100-day mortality)을 예측하기 위해 데이터 마이닝 기반 접근 방식을 이용하여 개발될 수 있다. 일본 레지스트리 데이터를 이용하여 급성 GVHD의 발병률(incidence)을 예측하는 기계 학습 알고리즘에 대한 또 다른 연구에서, 계산된 스코어가 급성 GVHD의 명확한 계층화(clear stratification)와 관련되는 것은 발견될 수 있다. 구체적으로, 급성 GVHD의 발병률이 낮으면 낮을수록 스코어가 낮을 수 있다.

[0007] 한 연구에서, 급성 백혈병 환자에서 동종 HCT 후 1년 재발률(1-year relapse)을 예측하기 위한 기계 학습 기반 모델은 개발될 수 있다. 환자 모집단, 엔드포인트(endpoints), 변수 선택의 기준은 연구마다 다르지만, 성능은 유사하고 이전에 보고된 이식 결과 예측 스코어에 비하여 약간 더 나은 것처럼 보일 수 있다.

[0008] 그러나, 동종 HCT 후 생존은 질병 재발 및 GVHD, 생착 실패(engraftment failure) 또는 비재발 사망률 증가로 이어질 수 있는 감염을 포함한 이식 관련 합병증과 같은 다수의 변수들에 따라 달라질 수 있다. 또한, 이러한 HCT 합병증은 공여자 또는 수여자 관련 팩터들, 공여자-수여자 관계, 컨디셔닝(conditioning) 등의 다양한 팩터들과 관련될 수 있다.

발명의 내용

해결하려는 과제

[0010] 일 실시예에 따른 예측 모델은, 다수의 피쳐들을 동시에 적용하고 잠재적인 편향들을 감소시키기 위해 기계 학습 기반의 알고리즘을 이용할 수 있다.

[0011] 기계 학습 기반의 접근 방식에 의한 피쳐들의 선택 및 선택된 피쳐들을 적용한 예측 모델을 구축을 통해 모델의 성능은 향상될 수 있고, 예기치 않은 편향들은 방지될 수 있다.

[0012] 일 실시예에 따른 예측 모델은, HCT 후 결과를 개선하기 위해 더 나은 이식 관련 팩터들(transplantation-related factors) 또는 공여자(donors)를 선택하는 측면에서 도움이 될 수 있다.

[0013] 기계 학습 기반의 접근 방식에 의하여 선택된 피쳐들에 기초하여 동종 HCT 후 혈액 악성 종양 환자의 장기 생존 예측 모델은 개발될 수 있고, 검증 세트에서 모델의 정확도는 검증될 수 있다. 그 이후에, 개발된 예측 모델을 이용하여 보다 더 적절한 이식 관련 팩터를 선택할 수 있다.

과제의 해결 수단

[0015] 일 실시예에 따른 프로세서에 의하여 수행되는 방법은, 공여자(donor)에 관한 제1 의료 피쳐들(medical features), 수여자(recipient)에 관한 제2 의료 피쳐들, 및 상기 공여자와 상기 수여자 간의 관계(relationship)에 관한 제3 의료 피쳐들을 가지는 제1 입력 데이터에 기계 학습 모델을 적용하여 동종 조혈 세

포 이식(allogeneic stem cell transplantation)의 수여자가 미리 정의된 기간 내 사망할 제1 가능성 점수를 획득하는 단계, 상기 제1 입력 데이터의 의료 피쳐들 중 적어도 일부의 값을 변경함으로써 획득된 제2 입력 데이터에 상기 기계 학습 모델을 적용하여 제2 가능성 점수 획득하는 단계, 및 상기 제1 가능성 점수 및 상기 제2 가능성 점수를 출력하는 단계를 포함할 수 있다.

- [0016] 일 실시예에 따른 방법은, 상기 제1 입력 데이터 및 상기 제2 입력 데이터의 의료 피쳐들의 값들 각각의 상기 제1 가능성 점수 및 제2 가능성 점수에 대한 영향력을 나타내는 피쳐 기여도(feature contribution)를 계산하는 단계 및 상기 계산된 피쳐 기여도를 출력하는 단계를 더 포함할 수 있다.
- [0017] 상기 피쳐 기여도를 출력하는 단계는, 상기 피쳐 기여도의 크기에 따라 정렬된 상기 제1 입력 데이터의 상기 의료 피쳐들의 상기 피쳐 기여도를 출력하는 단계 및 상기 피쳐 기여도의 크기에 따라 정렬된 상기 제2 입력 데이터의 상기 의료 피쳐들의 상기 피쳐 기여도를 출력하는 단계를 포함할 수 있다.
- [0018] 상기 제2 가능성 점수 획득하는 단계는, 다른 공여자에 대한 의료 피쳐 값들을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 상기 제1 의료 피쳐들 및 제3 의료 피쳐들 중 적어도 일부의 값을 상기 공여자에 대응하는 값에서 상기 다른 공여자에 대응하는 값으로 변경함으로써 상기 제2 입력 데이터를 획득하는 단계 및 상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득하는 단계를 포함할 수 있다.
- [0019] 상기 제2 가능성 점수 획득하는 단계는, 다른 수여자에 대한 의료 피쳐 값들을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 상기 제2 의료 피쳐들 및 상기 제3 의료 피쳐들 중 적어도 일부의 값을 상기 수여자에 대응하는 값에서 상기 다른 수여자에 대응하는 값으로 변경함으로써 상기 제2 입력 데이터를 획득하는 단계 및 상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득하는 단계를 포함할 수 있다.
- [0020] 상기 제2 가능성 점수 획득하는 단계는, 동종 조혈 세포 이식의 계획(planning)에 따라 조정된 의료 피쳐 값을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 조정가능한 의료 피쳐들 중 적어도 일부의 값을 상기 조정된 의료 피쳐 값으로 변경함으로써 상기 제2 입력 데이터를 획득하는 단계 및 상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득하는 단계를 포함할 수 있다.
- [0021] 일 실시예에 따른 방법은, 복수의 의료 피쳐들을 포함하는 임시 입력 데이터에 기초하여 트레이닝된 임시 기계 학습 모델을 이용하여 상기 복수의 의료 피쳐들 중 상기 기계 학습 모델의 입력 데이터의 의료 피쳐들을 결정하는 단계 및 상기 입력 데이터의 의료 피쳐들을 가지는 트레이닝 입력 데이터에 기초하여 트레이닝된 상기 기계 학습 모델을 획득하는 단계를 더 포함할 수 있다.
- [0022] 상기 기계 학습 모델의 입력 데이터의 의료 피쳐들을 결정하는 단계는, 상기 임시 입력 데이터에 기초하여 상기 임시 기계 학습 모델을 트레이닝시키는 단계, 상기 트레이닝된 임시 기계 학습 모델에 대한 상기 임시 입력 데이터의 복수의 의료 피쳐들의 피쳐 중요도(feature importance)에 기초하여 상기 임시 입력 데이터의 상기 복수의 의료 피쳐들 중 일부를 상기 임시 입력 데이터에서 제거하는 단계, 및 상기 임시 기계 학습 모델의 트레이닝 및 상기 임시 입력 데이터에 대한 의료 피쳐 제거를 반복하는 단계를 포함할 수 있다.
- [0023] 상기 임시 기계 학습 모델의 트레이닝 및 상기 임시 입력 데이터의 의료 피쳐 제거를 반복하는 단계는, 상기 임시 기계 학습 모델의 트레이닝 및 상기 임시 입력 데이터의 의료 피쳐 제거와 함께 상기 임시 기계 학습 모델의 정확도 계산을 반복하는 단계 및 상기 계산된 정확도에 기초하여 상기 임시 입력 데이터의 의료 피쳐들을 상기 기계 학습 모델의 상기 입력 데이터의 상기 의료 피쳐들로 결정하는 단계를 포함할 수 있다.
- [0024] 일 실시예에 따른 전자 장치는, 공여자(donor)에 관한 제1 의료 피쳐들(medical features), 수여자(recipient)에 관한 제2 의료 피쳐들, 및 상기 공여자와 상기 수여자 간의 관계(relationship)에 관한 제3 의료 피쳐들을 가지는 제1 입력 데이터에 기계 학습 모델을 적용하여 동종 조혈 세포 이식(allogeneic stem cell transplantation)의 수여자가 미리 정의된 기간 내 사망할 제1 가능성 점수를 획득하고, 상기 제1 입력 데이터의 의료 피쳐들 중 적어도 일부의 값을 변경함으로써 획득된 제2 입력 데이터에 상기 기계 학습 모델을 적용하여 제2 가능성 점수 획득하는 프로세서 및 상기 제1 가능성 점수 및 상기 제2 가능성 점수를 출력하는 디스플레이를 포함할 수 있다.
- [0025] 상기 프로세서는, 상기 제1 입력 데이터 및 상기 제2 입력 데이터의 의료 피쳐들의 값들 각각의 상기 제1 가능성 점수 및 제2 가능성 점수에 대한 영향력을 나타내는 피쳐 기여도(feature contribution)를 계산할 수 있고,

상기 디스플레이는, 상기 계산된 피쳐 기여도를 출력할 수 있다.

- [0026] 상기 프로세서는, 상기 피쳐 기여도의 크기에 따라 정렬된 상기 제1 입력 데이터의 상기 의료 피쳐들의 상기 피쳐 기여도를 출력하고, 상기 피쳐 기여도의 크기에 따라 정렬된 상기 제2 입력 데이터의 상기 의료 피쳐들의 상기 피쳐 기여도를 출력할 수 있다.
- [0027] 상기 프로세서는, 다른 공여자에 대한 의료 피쳐 값들을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 상기 제1 의료 피쳐들 및 제3 의료 피쳐들 중 적어도 일부의 값을 상기 공여자에 대응하는 값에서 상기 다른 공여자에 대응하는 값으로 변경함으로써 상기 제2 입력 데이터를 획득하고, 상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득할 수 있다.
- [0028] 상기 프로세서는, 다른 수여자에 대한 의료 피쳐 값들을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 상기 제2 의료 피쳐들 및 상기 제3 의료 피쳐들 중 적어도 일부의 값을 상기 수여자에 대응하는 값에서 상기 다른 수여자에 대응하는 값으로 변경함으로써 상기 제2 입력 데이터를 획득하고, 상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득할 수 있다.
- [0029] 상기 프로세서는, 동종 조혈 세포 이식의 계획(planning)에 따라 조정된 의료 피쳐 값을 더 획득한 경우에 응답하여, 상기 제1 입력 데이터의 조정가능한 의료 피쳐들 중 적어도 일부의 값을 상기 조정된 의료 피쳐 값으로 변경함으로써 상기 제2 입력 데이터를 획득하고, 상기 제2 입력 데이터에 상기 기계 학습 모델을 적용함으로써 상기 제2 가능성 점수를 획득할 수 있다.
- [0030] 상기 프로세서는, 복수의 의료 피쳐들을 포함하는 임시 입력 데이터에 기초하여 트레이닝된 임시 기계 학습 모델을 이용하여 상기 복수의 의료 피쳐들 중 상기 기계 학습 모델의 입력 데이터의 의료 피쳐들을 결정하고, 상기 입력 데이터의 의료 피쳐들을 가지는 트레이닝 입력 데이터에 기초하여 트레이닝된 상기 기계 학습 모델을 획득할 수 있다.
- [0031] 상기 프로세서는, 상기 임시 입력 데이터에 기초하여 상기 임시 기계 학습 모델을 트레이닝시키고, 상기 트레이닝된 임시 기계 학습 모델에 대한 상기 임시 입력 데이터의 복수의 의료 피쳐들의 피쳐 중요도(feature importance)에 기초하여 상기 임시 입력 데이터의 상기 복수의 의료 피쳐들 중 일부를 상기 임시 입력 데이터에서 제거하며, 상기 임시 기계 학습 모델의 트레이닝 및 상기 임시 입력 데이터에 대한 의료 피쳐 제거를 반복할 수 있다.
- [0032] 상기 프로세서는, 상기 임시 기계 학습 모델의 트레이닝 및 상기 임시 입력 데이터의 의료 피쳐 제거와 함께 상기 임시 기계 학습 모델의 정확도 계산을 반복하고, 상기 계산된 정확도에 기초하여 상기 임시 입력 데이터의 의료 피쳐들을 상기 기계 학습 모델의 상기 입력 데이터의 상기 의료 피쳐들로 결정할 수 있다.

발명의 효과

- [0034] 일 실시예에 따른 예측 모델은 0.03 표준편차와 함께 0.788의 ROC 커브 아래의 평균 면적(mean area under the receiver operating characteristic curve)의 성능을 가질 수 있다. 이식 후 생존의 확률을 예측하는 위험 점수(risk score)는 랜덤하게 선택된 294명의 환자들을 통해 개발될 수 있고, 위험 점수에 따른 전체 사망(overall death), 비재발 사망률(non-relapse mortality), 및 재발 발생(relapse incidence)의 추정된 예측 및 관찰 위험은 서로 일치(agreement)를 가질 수 있다. 또한 계산된 점수는 각 변수의 중요도를 가시화하여 다른 이식 관련 팩터에 따른 생존을 예측하는 것의 가능성을 나타낼 수 있다.
- [0035] 혈액암 환자에서 동종 조혈 세포 이식 후 미리 정의된 기간 내 사망할 확률의 정량화를 위하여, 기계 학습 모델 및 가능성 점수는 제시될 수 있다. 가능성 점수는 전체 생존, 비재발 사망률(NRM), 및 재발의 발생률과 좋은 상관관계를 가질 수 있다. 일 실시예에 따른 기계 학습 모델은 동종 조혈 세포 이식의 계획(planning)을 선택하는 데 도움될 수 있다. 또한, 일 실시예에 따른 기계 학습 모델은 수여자 후보 또는 공여자 후보를 선택하기 위한 동종 조혈 세포 이식 케이스 별 접근법을 제공할 수 있다.

도면의 간단한 설명

- [0037] 도 1은 일 실시예에 따른 기계 학습 모델을 통한 동종 조혈 세포 이식을 받은 수여자가 장기 생존할 가능성 점

수를 출력하는 전자 장치를 나타낸다.

도 2는 일 실시예에 따라 동종 조혈 세포 이식 케이스의 수여자가 미리 정의된 기간 내 사망할 가능성 점수를 출력하는 전자 장치의 동작을 나타낸다.

도 3은 일 실시예에 따른 기계 학습 모델의 입력 데이터의 의료 피처들의 선택을 나타낼 수 있다.

도 4는 일 실시예에 따른 디스플레이의 가능성 점수 및 피처 기여도의 출력을 나타낸다.

도 5는 일 실시예에 따른 기계 학습 모델의 성능을 10겹 교차 검증을 이용한 평가를 나타낸다.

도 6은 일 실시예에 따른 기계 학습 모델의 가능성 점수에 기초하여 분할된 환자 클래스에 따른 전체 생존, 재발의 발생률, 및 비재발 사망률을 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0038] 실시예들에 대한 특정한 구조적 또는 기능적 설명들은 단지 예시를 위한 목적으로 개시된 것으로서, 다양한 형태로 변경되어 구현될 수 있다. 따라서, 실제 구현되는 형태는 개시된 특정 실시예로만 한정되는 것이 아니며, 본 명세서의 범위는 실시예들로 설명한 기술적 사상에 포함되는 변경, 균등물, 또는 대체물을 포함한다.
- [0039] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 이런 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 해석되어야 한다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.
- [0040] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다.
- [0041] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 설명된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로써 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0042] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 해당 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0043] 이하, 실시예들을 첨부된 도면들을 참조하여 상세하게 설명한다. 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조 부호를 부여하고, 이에 대한 중복되는 설명은 생략하기로 한다.
- [0044] 도 1은 일 실시예에 따른 기계 학습 모델을 통한 동종 조혈 세포 이식을 받은 수여자가 장기 생존할 가능성 점수를 출력하는 전자 장치를 나타낸다.
- [0045] 전자 장치(100)는 프로세서(110) 및 디스플레이(120)를 포함할 수 있다.
- [0046] 프로세서(110)는, 입력 데이터(input data)에 기계 학습 모델을 적용함으로써 가능성 점수(possibility score)를 출력할 수 있다. 입력 데이터는 의료 피처들(medical features)을 포함하는 의료 데이터(medical data)를 포함할 수 있다.
- [0047] 의료 피처들은 공여자(donor)에 관한 제1 의료 피처들, 수여자(recipient)에 관한 제2 의료 피처들, 및 공여자 및 수여자의 관계(relationship)에 관한 제3 의료 피처들을 포함할 수 있다. 의료 피처는, 동종 조혈 세포 이식 케이스의 의료 상황과 관련된 정보를 분류한 개별 항목 및/또는 카테고리를 나타낼 수 있다.
- [0048] 동종 조혈 세포 이식 케이스는, 동종 조혈 세포 이식의 하나의 사례를 나타내는 것으로, 공여자에 관한 의료 정보, 수여자에 관한 의료 정보, 및 동종 조혈 세포 이식의 계획의 조합에 대응할 수 있다. 예를 들어, 동일한 공여자의 동종 조혈 세포 이식 케이스들의 수여자가 서로 다르면 서로 다른 이식 케이스들을 나타낼 수 있고, 동종 조혈 세포 이식 케이스들의 공여자 및 수여자가 동일한 사람인 경우에도 공여자 및 수여자 중 적어도 한 사람의 의료 정보(예: 나이, 질병 특성 등)가 다른 경우에는 서로 다른 이식 케이스들을 나타낼 수 있다. 일

실시예에 따른 입력 데이터의 의료 피처의 선택은 도 3에서 후술한다.

- [0049] 공여자에 관한 제1 의료 피처는, 공여자 특성을 나타내는 의료 피처로서, 공여자의 HLA-A 유형, 공여자의 HLA-B 유형, 공여자의 HLA-C 유형, 공여자의 HLA-DR 유형, 및 공여자의 CMV IgG 유형을 포함할 수 있다. 수여자에 관한 제2 의료 피처는, 수여자 특성을 나타내는 의료 피처로서, 수여자의 HLA-A 유형, 수여자의 HLA-B 유형, 수여자의 HLA-C 유형, 수여자의 HLA-DR 유형, 수여자의 CMV IgG 유형, 수여자의 간의 HCT-CI 스코어, HCT-CI 스코어, 수여자의 진단 질병, HCT에서 질병 상태, 진단 시 백혈구, HCT 전의 적혈구 수혈(red blood cell transfusion), 및 HCT 전의 혈소판 수혈(platelet transfusion)을 포함할 수 있다. 공여자 및 수여자의 관계에 관한 제3의료 피처는, 공여자 및 수여자 특성 간의 매칭된 정도를 나타내는 의료 피처로서, HLA-A, B, C, DR 미스매칭된 유전자좌들(mismatched loci)의 개수, 및 HLA-DR 미스매칭된 유전자좌들의 개수를 포함할 수 있다.
- [0050] 다만, 이에 한정하는 것은 아니고, 예를 들어, 의료 피처는 수여자 특성 및 공여자 특성과 함께, 질병 특징, 인간 백혈구 항원(human leukocyte antigen; HLA) 유형, 이식 정보(graft information), 투여된 컨디셔닝 약물(administered medications of conditioning), GVHD 예방, 및 지지적 치료(supportive care), 및 기타 실험실 데이터(other laboratory data)를 포함할 수 있다.
- [0051] 가능성 점수는 공여자로부터 동종 조혈 세포 이식을 받은 수여자가 동종 조혈 세포 이식 후 미리 정의된 기간 내(예: 5년) 사망할 가능성을 나타낼 수 있다.
- [0052] 예를 들어, 프로세서(110)는 제1 사람의 동종 조혈 세포를 제2 사람에게 이식하는 경우에 대한 가능성 점수를 출력할 수 있다. 입력 데이터는 공여자인 제1 사람에 대응하는 값의 제1 의료 피처들, 수여자인 제2 사람에 대응하는 값의 제2 의료 피처들, 및 공여자인 제1 사람 및 수여자인 제2 사람 간의 관계에 대응하는 값의 제3 의료 피처들을 포함할 수 있다. 가능성 점수는 제1 사람의 동종 조혈 세포를 제2 사람에게 이식한 후 제2 사람이 5년 이내 사망할 가능성을 나타낼 수 있다. 입력 데이터는 하나의 동종 조혈 세포 이식 케이스에 대응할 수 있다. 예를 들어, 입력 데이터는 동종 조혈 세포 이식 케이스에 따른 값들을 갖는 의료 피처들을 포함할 수 있다.
- [0053] 후술하겠으나, 하나의 공여자에 복수의 수여자가 매핑되거나 복수의 공여자에 하나의 수여자가 매핑되는 경우, 공여자 및 수여자 조합 각각에 대하여 가능성 점수들을 계산하여 비교할 수 있다. 또는, 동일한 공여자 및 수여자에 대하여, 조혈 세포 이식의 계획에 따라 조정가능한 피처의 여러가지 값들 각각에 대하여 가능성 점수들을 계산하여 비교할 수 있다. 가능성 점수들의 계산 및 비교는 도 4에서 후술한다.
- [0054] 디스플레이(120)는 가능성 점수를 출력할 수 있다. 디스플레이(120)는 가능성 점수와 함께 기계 학습 모델(112)에 입력 데이터의 의료 피처들 중 적어도 일부의 피처 기여도를 함께 출력할 수 있다. 디스플레이의 가능성 점수 및 피처 기여도 출력은 도 4에서 후술한다.
- [0055] 도 2는 일 실시예에 따라 동종 조혈 세포 이식 케이스의 수여자가 미리 정의된 기간 내 사망할 가능성 점수를 출력하는 전자 장치의 동작을 나타낸다.
- [0056] 단계(210)에서, 프로세서는 임시 입력 데이터의 복수의 의료 피처들 중 입력 데이터의 의료 피처들을 결정할 수 있다. 임시 입력 데이터는 수집된 의료 피처들 중 적어도 일부를 포함하는 데이터를 나타낼 수 있다. 임시 입력 데이터의 복수의 의료 피처들 중 일부를 입력 데이터의 의료 피처들로 선택할 수 있다. 프로세서는 임시 입력 데이터에 기초하여 트레이닝된 임시 기계 학습 모델을 이용하여 입력 데이터의 의료 피처들을 선택할 수 있다. 입력 데이터의 의료 피처들의 선택은 도 3에서 후술한다.
- [0057] 기계 학습 모델의 트레이닝 데이터 세트는, 선택된 의료 피처들을 포함하는 트레이닝 입력 데이터 및 트레이닝 입력 데이터에 매핑된 참값 클래스를 포함할 수 있다. 트레이닝 입력 데이터는, 선택된 입력 데이터의 의료 피처들을 포함하는 데이터로서, 하나의 동종 조혈 세포 이식 케이스를 나타낼 수 있다. 참값 클래스는, 동종 조혈 세포 이식 케이스의 수여자가 이식 이후 미리 정의된 기간 내에 사망할지 여부를 나타낼 수 있다. 예시적으로, 참값 클래스는 동종 조혈 세포 이식을 받은 수여자가 이식 이후 5년을 초과하는 기간 동안 생존한 경우에 생존 클래스(예: 0의 참값 가능성 점수)로, 이식 이후 5년 내의 기간 동안 사망한 경우 사망 클래스(예: 1의 참값 가능성 점수)로 라벨링될 수 있다.
- [0058] 단계(220)에서, 프로세서는 트레이닝 입력 데이터에 기초하여 트레이닝된 기계 학습 모델을 획득할 수 있다. 기계 학습 모델은 선택된 의료 피처들을 가지는 입력 데이터에 적용됨으로써 가능성 점수를 출력하도록 트레이닝될 수 있다.

- [0059] 예를 들어, 프로세서는 트레이닝 입력 데이터에 기계 학습 모델을 적용하여 가능성 점수를 획득할 수 있다. 프로세서는 가능성 점수 및 참값 클래스에 기초하여 기계 학습 모델의 목적함수 값을 계산할 수 있다. 프로세서는 계산된 목적함수 값에 기초하여 가능성 점수 및 참값 클래스가 수렴하도록 기계 학습 모델의 파라미터를 업데이트할 수 있다. 프로세서는 목적함수 값의 계산 및 파라미터의 업데이트를 반복함으로써, 트레이닝된 기계 학습 모델을 획득할 수 있다.
- [0060] 단계(230)에서, 프로세서는 제1 입력 데이터에 기계 학습 모델을 적용하여 제1 가능성 점수를 획득할 수 있다. 제1 입력 데이터는 제1 동종 조혈 세포 이식 케이스를 나타낼 수 있다. 제1 가능성 점수는, 제1 동종 조혈 세포 이식 케이스에서, 수여자가 이식 후 미리 정의된 기간 내에 사망할 가능성을 나타낼 수 있다. 제1 입력 데이터는 공여자 및 수여자 중 적어도 하나에 대응하는 값의 의료 피쳐들을 가질 수 있다. 입력 데이터의 의료 피쳐들은 공여자에 관한 제1 의료 피쳐들, 수여자에 관한 제2 의료 피쳐들, 및 공여자 및 수여자의 관계에 관한 제3 의료 피쳐들을 포함할 수 있다.
- [0061] 단계(240)에서, 프로세서는 제2 입력 데이터에 기계 학습 모델을 적용하여 제2 가능성 점수를 획득할 수 있다. 제2 입력 데이터는 제1 동종 조혈 세포 케이스와 다른 제2 동종 조혈 세포 이식 케이스를 나타낼 수 있다. 제2 가능성 점수는, 제2 동종 조혈 세포 이식 케이스에서, 수여자가 이식 후 미리 정의된 기간 내에 사망할 가능성을 나타낼 수 있다. 프로세서는 제1 입력 데이터의 적어도 일부의 값을 변경함으로써 제2 입력 데이터를 획득할 수 있다.
- [0062] 예를 들어, 프로세서는 하나의 수여자에 매핑된 서로 다른 공여자들의 동종 조혈 세포 이식 케이스들을 비교할 수 있다. 프로세서는 다른 공여자에 대한 의료 피쳐 값들을 더 획득할 수 있다. 프로세서는 다른 공여자에 대한 의료 피쳐 값들을 더 획득한 경우에 응답하여, 제2 입력 데이터를 획득할 수 있다. 제2 입력 데이터는 다른 공여자로부터 수여자로의 동종 조혈 세포 이식 케이스를 나타낼 수 있다. 프로세서는, 제1 입력 데이터의 의료 피쳐들 중 공여자와 연관된 의료 피쳐들(예: 제1 의료 피쳐들 및 제3 의료 피쳐들) 중 적어도 일부의 값을 변경할 수 있다. 프로세서는 제1 입력 데이터 중 적어도 일부의 값을 공여자에 대응하는 값에서 다른 공여자에 대응하는 값으로 변경할 수 있다.
- [0063] 다른 예를 들어, 프로세서는 하나의 공여자에 매핑된 서로 다른 수여자들의 동종 조혈 세포 이식 케이스들을 비교할 수 있다. 프로세서는 다른 수여자에 대한 의료 피쳐 값들을 더 획득할 수 있다. 프로세서는 다른 수여자에 대한 의료 피쳐 값들을 더 획득한 경우에 응답하여, 제2 입력 데이터를 획득할 수 있다. 제2 입력 데이터는 공여자로부터 다른 수여자로의 동종 조혈 세포 이식 케이스를 나타낼 수 있다. 프로세서는, 제1 입력 데이터의 의료 피쳐들 중 수여자와 연관된 의료 피쳐들(예: 제2 의료 피쳐들 및 제3 의료 피쳐들) 중 적어도 일부의 값을 변경할 수 있다. 프로세서는 제1 입력 데이터 중 적어도 일부의 값을 수여자에 대응하는 값에서 다른 수여자에 대응하는 값으로 변경할 수 있다.
- [0064] 또 다른 예를 들어, 프로세서는 공여자 및 수여자의 동종 조혈 세포 이식의 서로 다른 계획들에 따른 동종 조혈 세포 이식 케이스들을 비교할 수 있다. 입력 데이터의 의료 피쳐들 중 적어도 일부는 동종 조혈 세포 이식의 계획에 따라 조정가능할 수 있다. 예를 들어, 입력 데이터는 하나의 공여자로부터 하나의 수여자에게 동종 조혈 세포를 이식하는 경우에도, 동종 조혈 세포 이식의 계획에 따라 복수의 값들 중 하나의 값을 가지는 의료 피쳐들을 포함할 수 있다. 참고로, 동종 조혈 세포 이식의 계획에 따라 조정가능한 의료 피쳐들은, 공여자에 관한 의료 피쳐들, 수여자에 관한 의료 피쳐들, 및 공여자 및 수여자의 관계에 관한 의료 피쳐들의 적어도 일부를 포함할 수 있다. 예시적으로, 동종 조혈 세포 이식의 계획에 따라 조정가능한 의료 피쳐는 전처치 요법(Conditioning regimen)의 강도 및 약물의 종류, GVHD 예방약물의 종류 및 용량, 이식편의 종류(예: 말초혈액 또는 골수), 및 환자-공여자 관계(예: 형제, HLA 반일치 가족, 또는 타인)을 포함할 수 있다. 전처치 요법은, 이식 전 생착을 위한 공간확보, 면역조절 및 항암효과 등을 위해 환자에게 투여되는 약제를 나타낼 수 있다. 프로세서는, 동종 조혈 세포의 이식의 계획에 따라 조정가능한 의료 피쳐 값들을 더 획득할 수 있다. 프로세서는, 조정가능한 의료 피쳐들 중 적어도 일부의 값을 변경할 수 있다. 프로세서는 제1 입력 데이터 중 적어도 일부의 값을 제1 입력 데이터의 동종 조혈 세포 이식의 계획에 대응하는 값에서 다른 계획에 대응하는 값으로 변경할 수 있다.
- [0065] 단계(250)에서, 디스플레이는 제1 가능성 점수 및 제2 가능성 점수를 출력할 수 있다. 디스플레이는 가능성 점수와 함께, 입력 데이터의 의료 피쳐들의 영향력을 나타내는 피쳐 기여도를 함께 출력할 수 있다. 가능성 점수 및 피쳐 기여도의 출력은 도 4에서 후술한다.

- [0066] 도 3은 일 실시예에 따른 기계 학습 모델의 입력 데이터의 의료 피쳐들의 선택을 나타낼 수 있다.
- [0067] 단계(310)에서, 프로세서는 임시 입력 데이터에 기초하여 임시 기계 학습 모델을 트레이닝시킬 수 있다.
- [0068] 단계(320)에서, 프로세서는 피쳐 중요도(feature importance)에 기초하여 임시 입력 데이터의 복수의 의료 피쳐들 중 일부를 임시 입력 데이터에서 제거할 수 있다. 피쳐 중요도는 트레이닝된 임시 기계 학습 모델에 대한 임시 입력 데이터의 각 의료 피쳐의 중요성을 나타낼 수 있다. 피쳐 중요도는 임시 입력 데이터의 의료 피쳐들 각각에 대하여 계산될 수 있다. 프로세서는 피쳐 중요도에 기초하여 의료 피쳐들 중 일부를 임시 입력 데이터에서 제거할 수 있다. 예를 들어, 프로세서는 임시 입력 데이터의 의료 피쳐들 중 최소 피쳐 중요도를 가지는 의료 피쳐를 제거할 수 있다. 다만, 이에 한정하는 것은 아니고, 프로세서는 임시 입력 데이터의 의료 피쳐들 중 임계 피쳐 중요도 이하의 피쳐 중요도를 가지는 의료 피쳐를 제거할 수도 있다.
- [0069] 프로세서는 임시 기계 학습 모델의 정확도를 계산할 수 있다. 프로세서는 가능성 점수를 임계 점수와 비교함으로써 수여자가 동종 조혈 세포 이식 후 미리 정의된 기간 내에 사망할지 여부를 지시하는 클래스를 결정할 수 있다. 예를 들어, 프로세서는 가능성 점수가 임계 점수 이상인 경우에 응답하여 사망 클래스로 결정하고, 가능성 점수가 임계 점수 미만인 경우에 응답하여 생존 클래스로 결정할 수 있다. 프로세서는 임시 기계 학습 모델에 기초하여 결정된 클래스와 트레이닝 데이터 세트의 참값 클래스를 비교함으로써 정확도를 계산할 수 있다.
- [0070] 단계(330)에서, 프로세서는 임시 기계 학습 모델의 트레이닝 및 임시 입력 데이터에 대한 의료 피쳐 제거를 반복할 수 있다.
- [0071] 초기의 임시 입력 데이터는 수집된 의료 피쳐들을 모두 포함할 수 있다. 프로세서는 피쳐 중요도에 기초하여 의료 피쳐들 중 일부를 임시 입력 데이터에서 제거할 수 있다. 프로세서는 일부 의료 피쳐가 제거된 임시 입력 데이터를 임시 입력 데이터로 업데이트할 수 있다.
- [0072] 프로세서는 임시 기계 학습 모델의 트레이닝 및 임시 입력 데이터에 대한 의료 피쳐 제거와 함께, 임시 기계 학습 모델의 정확도 계산을 반복할 수 있다. 프로세서는 임시 기계 학습 모델의 정확도에 기초하여 임시 기계 학습 모델의 트레이닝 및 임시 입력 데이터에 대한 의료 피쳐 제거를 중단할 수 있다. 프로세서는 임시 기계 학습 모델 및 임시 입력 데이터에 기초하여 기계 학습 모델 및 입력 데이터의 의료 피쳐들을 획득할 수 있다. 예를 들어, 프로세서는 임시 기계 학습 모델의 정확도가 이전 반복에서의 임시 기계 학습 모델보다 작은 경우에 응답하여, 임시 기계 학습 모델의 트레이닝 및 임시 입력 데이터에 대한 의료 피쳐 제거를 중단할 수 있다. 예를 들어, 프로세서는 이전 반복에서의 임시 기계 학습 모델을 기계 학습 모델로, 이전 반복에서의 임시 입력 데이터의 의료 피쳐들을 입력 데이터의 의료 피쳐들로 획득할 수 있다.
- [0073] 예를 들어, 프로세서는 제1 차원의 제1 임시 입력 데이터에 기초하여 제1 임시 기계 학습 모델을 트레이닝시킬 수 있다. 제1 차원은 제1 임시 입력 데이터의 원소(element)의 개수로서, 제1 임시 입력 데이터의 의료 피쳐들의 개수에 대응할 수 있다. 예를 들어, 제1 임시 입력 데이터가 100개의 의료 피쳐들을 가지는 경우, 제1 임시 입력 데이터의 차원은 100일 수 있다. 제1 임시 기계 학습 모델은 제1 임시 입력 데이터에 적용됨으로써 가능성 점수를 출력하는 모델을 나타낼 수 있다.
- [0074] 프로세서는 제1 임시 입력 데이터의 의료 피쳐들의 피쳐 중요도에 기초하여, 제1 임시 입력 데이터에서 의료 피쳐들 중 일부를 제거할 수 있다. 프로세서는 제1 임시 입력 데이터에서 일부 의료 피쳐를 제거함으로써 제2 임시 입력 데이터를 획득할 수 있다. 제2 임시 입력 데이터는 제1 임시 입력 데이터보다 적은 개수의 의료 피쳐들을 가질 수 있다. 제2 임시 입력 데이터는 제1 임시 입력 데이터보다 더 작은 차원을 가질 수 있다.
- [0075] 프로세서는 제2 임시 입력 데이터에 기초하여 제2 임시 기계 학습 모델을 트레이닝시킬 수 있다. 제2 임시 기계 학습 모델은 제2 차원의 제2 임시 입력 데이터에 적용됨으로써 가능성 점수를 출력하는 모델을 나타낼 수 있다. 제2 임시 기계 학습 모델은 제1 임시 기계 학습 모델보다 더 작은 차원의 입력 데이터에 적용됨으로써 가능성 점수를 출력할 수 있다.
- [0076] 프로세서는 임시 기계 학습 모델의 트레이닝 및 임시 입력 데이터에 대한 의료 피쳐 제거를 반복함으로써, 더 작은 차원의 임시 입력 데이터에 적용되는 임시 기계 학습 모델을 획득할 수 있다. 일 실시예에 따른 기계 학습 모델은 의료 피쳐 제거를 통해 수집된 의료 피쳐들을 모두 포함하는 입력 데이터에 적용되는 임시 기계 학습 모델보다 더 큰 정확도를 가질 수 있다. 기계 학습 모델의 입력 데이터가 복수의 의료 피쳐들을 모두를 포함하는 경우보다 복수의 의료 피쳐들 중 선택된 일부의 의료 피쳐들을 포함하는 경우가 더 피쳐 개수 대비 효율적인 성능을 갖는 기계 학습 모델을 획득할 수 있다.

[0077] 더 작은 차원의 입력 데이터에 적용되는 기계 학습 모델은 수집된 많은 의료 피쳐들 중에서 가능성 점수에 중요한 의료 피쳐를 암시할 수 있다. 따라서, 작은 차원의 입력 데이터에 적용되는 기계 학습 모델은 큰 차원의 임시 입력 데이터에 적용되는 임시 기계 학습 모델보다 더 설명 가능할(explainable) 수 있다.

[0078] 프로세서는 더 작은 차원의 입력 데이터에 기계 학습 모델을 적용하는 경우, 더 빠르게 실행될 수 있다. 기계 학습 모델이 더 적은 차원의 입력 데이터에 적용되는 경우, 수여자의 사망 가능성을 예측하기 위한 동종 조혈 세포 이식 케이스에 대한 의료 데이터 중 적어도 일부가 누락될 가능성이 감소할 수 있다.

[0079] 또한, 입력 데이터의 의료 피쳐는 복수의 의료 피쳐들 중에서 전문가(예: 의료인 등)에 의하여 수집된 의료 피쳐들 중에서 선택될 수도 있으나, 일 실시예에 따른 기계 학습 모델의 입력 데이터의 의료 피쳐들은 피쳐 중요도에 기초하여 선택될 수 있다. 전문가에 의하여 선택된 의료 피쳐들보다 기계 학습 모델에 대한 피쳐 중요도에 기초하여 의료 피쳐들은 보다 더 기계 학습 모델의 성능을 향상시킬 수 있다.

[0080] 도 4는 일 실시예에 따른 디스플레이의 가능성 점수 및 피쳐 기여도의 출력을 나타낸다.

[0081] 프로세서는 가능성 점수에 대한 의료 피쳐들의 값들 각각의 피쳐 기여도(feature contribution)를 계산할 수 있다. 피쳐 기여도는, 의료 피쳐들 각각의 값이 가능성 점수에 미치는 영향력을 나타낼 수 있다. 예를 들어, 프로세서는 제1 입력 데이터에 기계 학습 모델을 적용함으로써 제1 가능성 점수를 획득할 수 있다. 제1 입력 데이터의 의료 피쳐들의 값들 각각은 제1 가능성 점수에 대하여 서로 다른 정도의 영향력을 가질 수 있다. 프로세서는 제1 입력 데이터의 의료 피쳐들의 값들 각각에 대하여 제1 가능성 점수에 대한 피쳐 기여도를 계산할 수 있다. 또한, 프로세서는 제2 입력 데이터에 기계 학습 모델을 적용함으로써 제2 가능성 점수를 획득할 수 있다. 제2 입력 데이터의 의료 피쳐들의 값들 각각은 제2 가능성 점수에 대한 서로 다른 정도의 영향력을 가질 수 있다. 프로세서는 제2 입력 데이터의 의료 피쳐들의 값들 각각에 대하여 제2 가능성 점수에 대한 피쳐 기여도를 계산할 수 있다.

[0082] 일 실시예에 따른 프로세서는 샐플리 값(Shapley value)을 통해 피쳐 기여도를 계산할 수 있다. 샐플리 값은, 게임 이론으로부터 정의된 값으로서, 가능한 모든 의료 피쳐들의 값들의 조합들에 대해 하나의 의료 피쳐의 값에 따른 평균적인 변화를 통해 계산될 수 있다. 샐플리 값은 기계 학습 모델의 출력인 가능성 점수에 대한 설명을 제공할 수 있다. 가능성 점수를 출력하는 모델($f(x)$)이 주어지면, 다음 방정식을 이용하여 샐플리 값은 계산될 수 있다:

$$\varphi_i(f, x) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} [f_x(S \cup \{i\}) - f_x(S)]$$

[0083]

[0084] 여기서, n 은 입력 데이터의 의료 피쳐들의 총 개수를 나타낼 수 있고, 합계(summation)는 의료 피쳐 i 를 포함하지 않는 N 의 모든 부분집합들 S 에 걸쳐 확장될 수 있다. 예시적으로, 통합 프레임워크 SHAP(shapley additive explanation)는 샐플리 값을 이용하여 설명 가능한 기계 학습 모델을 구현하기 위하여 이용될 수 있다.

[0085] 디스플레이는 프로세서에 의하여 계산된 피쳐 기여도를 출력할 수 있다. 예를 들어, 디스플레이는 제1 가능성 점수에 대한 제1 입력 데이터의 의료 피쳐들의 값들의 피쳐 기여도를 출력할 수 있다. 디스플레이는 제2 가능성 점수에 대한 제2 입력 데이터의 의료 피쳐들의 값들의 피쳐 기여도를 출력할 수 있다.

[0086] 프로세서는 가능성 점수에 대한 피쳐 기여도들 중 적어도 일부를 크기에 따라 정렬할 수 있다. 예를 들어, 프로세서는 제1 가능성 점수에 대한 제1 입력 데이터의 의료 피쳐들의 값들의 피쳐 기여도들 중 적어도 일부를 크기에 따라 정렬할 수 있다. 프로세서는 제2 가능성 점수에 대한 제2 입력 데이터의 의료 피쳐들의 값들의 피쳐 기여도들 중 적어도 일부를 크기에 따라 정렬할 수 있다. 예를 들어, 프로세서는 피쳐 기여도들 중 적어도 일부를 내림차순(또는 오름차순)으로 정렬할 수 있다. 디스플레이는 정렬된 피쳐 기여도를 출력할 수 있다.

[0087] 일 실시예에 따른 디스플레이는 가능성 점수에 대한 의료 피쳐들의 값들의 피쳐 기여도를 그래프로 출력할 수 있다. 디스플레이는 그래프의 일축에 따라 의료 피쳐들을 출력하고, 그래프의 타축에 따라 가능성 점수 및 피쳐 기여도를 출력할 수 있다. 후술하겠으나, 그래프의 라인은 가능성 점수 및 피쳐 기여도를 나타낼 수 있다.

[0088] 디스플레이는 가능성 점수에 대한 그래프의 일축에 따라 의료 피쳐들을 출력할 수 있다. 예를 들어, 디스플레이는 제1 가능성 점수에 대한 그래프(410)의 일축을 따라 의료 피쳐들(412)을 출력할 수 있다. 디스플레이는 제2 가능성 점수에 대한 그래프(420)의 일축을 따라 의료 피쳐들(422)을 출력할 수 있다. 프로세서는 피쳐 기여도의 크기에 따라 의료 피쳐들을 내림차순으로 정렬할 수 있다. 디스플레이는 정렬된 의료 피쳐들을 출력할

수 있다. 예를 들어, 프로세서는 제1 가능성 점수에 대한 의료 피처(412a)보다 의료 피처(412b)의 피처 기여도가 큰 경우, 그래프(410)에서 의료 피처(412a)를 의료 피처(412b)보다 선행 위치에 배치할 수 있다.

[0089] 디스플레이는 그래프의 타측에 따라 가능성 점수 및 피처 기여도를 출력할 수 있다. 예를 들어, 디스플레이는 가능성 점수 및 피처 기여도를 나타내는 그래프의 라인을 출력할 수 있다. 그래프(410)의 라인(415)은 그래프의 타측 상의 제1 가능성 점수에 대응하는 지점(411)에서부터 시작될 수 있다. 그래프의 라인은 부분 라인들을 포함할 수 있다. 각 부분 라인은 의료 피처에 대응할 수 있다. 부분 라인(415a)은 해당 의료 피처(412a)에 대응하는 영역에 배치될 수 있다. 부분 라인(415a)은 제1 가능성 점수에 대한 의료 피처(412a)의 값(413a)의 피처 기여도에 따라 결정될 수 있다. 부분 라인(415a)은, 피처 기여도의 절댓값의 크기에 비례하는 타측 상의 길이(414a)를 가질 수 있다. 부분 라인(415a)은, 하나의 의료 피처에 대응하는 일측 상의 길이를 가질 수 있다. 부분 라인은 피처 기여도가 양수인 경우 부분 라인 상의 지점이 타측에 가까워짐에 따라 타측의 양의 방향으로 향하는 선분일 수 있다. 예를 들어, 의료 피처(412b)의 피처 기여도가 음수인 경우, 부분 라인(415b)은 타측에 가까워짐에 따라 타측의 양의 방향으로 향하는 선분일 수 있다. 부분 라인은 피처 기여도가 음수인 경우 부분 라인 상의 지점이 타측에 가까워짐에 따라 타측의 음의 방향으로 향하는 선분일 수 있다. 예를 들어, 의료 피처(412a)의 피처 기여도가 음수인 경우, 부분 라인(415a)은 타측에 가까워짐에 따라 타측의 음의 방향으로 향하는 선분일 수 있다.

[0090] 디스플레이는 그래프(410)에서 의료 피처들의 값들(413)을 출력할 수 있다. 디스플레이는 의료 피처들 각각에 대응하는 영역에 해당 의료 피처의 값을 출력할 수 있다. 디스플레이는 제1 가능성 점수에 대한 그래프(410)에서, 의료 피처(412a)에 대응하는 영역에 의료 피처(412a)의 값(413a)을 출력할 수 있다. 디스플레이는 제2 가능성 점수에 대한 그래프(420)에서, 의료 피처(422a)에 대응하는 영역에 의료 피처(422a)의 값(423a)을 출력할 수 있다. 예를 들어, 디스플레이는 그래프의 라인(예: 그래프(410)의 라인(415), 및 그래프(420)의 라인(425))의 주변 영역 중 의료 피처들 각각에 대응하는 영역에 해당 의료 피처의 값을 표시할 수 있다.

[0091] 디스플레이는 제1 가능성 점수에 대한 피처 기여도 및 제2 가능성 점수에 대한 피처 기여도를 출력할 수 있다. 디스플레이는 복수의 동종 조혈 세포 이식 케이스들을 비교하기 위한 비교 인터페이스를 제공할 수 있다. 예를 들어, 제1 입력 데이터는 제1 동종 조혈 세포 이식 케이스를 나타낼 수 있다. 제2 입력 데이터는 제2 동종 조혈 세포 이식 케이스를 나타낼 수 있다. 디스플레이는 제1 가능성 점수에 대한 피처 기여도를 제1 영역에 출력할 수 있다. 디스플레이는 제1 영역에 대응하는 제2 영역에 제2 가능성 점수에 대한 피처 기여도를 출력할 수 있다. 예를 들어, 그래프(410)의 일측 및 그래프(420)의 일측은 나란하게 배치될 수 있다. 다른 예를 들어, 그래프(410)의 타측 및 그래프(420)의 타측은 나란하게 배치될 수 있다.

[0092] 참고로, 복수의 공여자 후보들이 존재할 때, 가능성 점수는 가장 적합한 공여자를 선택하는 데 이용될 수 있다. 예를 들어, 가능성 점수는 나이가 더 적은 HLA-동일 단배체(HLA-haploidentical) 공여자 및 나이가 더 많은 일치 형제(matched sibling) 공여자 중 하나의 공여자를 선택하는 데 도움이 될 수 있다. 추가적으로, 가능성 점수에 대한 각 의료 피처의 영향력을 나타내는 피처 기여도는 계산될 수 있다. 피처 기여도는, 예시적으로 새플리 값을 포함할 수 있고, 각 의료 피처의 영향력을 시각화할 수 있다.

[0093] 도 4에서 나타난 바와 같이, 프로세서는 서로 다른 값을 나타내는 각 공여자에 대응하는 제1 입력 데이터 및 제2 입력 데이터에 기초하여 2개의 가능성 점수들을 획득할 수 있다. 이식 전 질병 상태(도 4의 그래프(410) 및 그래프(420)에서 stbmt로 표시됨)를 나타내는 의료 피처는 의료 피처들 중에서 가능성 점수를 계산하는 데 가장 영향력이 높은 의료 피처로 나타날 수 있다. 제2 가능성 점수가 제1 가능성 점수보다 낮기 때문에, 제1 입력 데이터의 공여자보다 제2 입력 데이터의 공여자가 선호될 수 있다.

[0094] 표 1은 일 실시예에 따른 데이터 세트에 포함된 환자들 및 공여자들의 특성들을 나타낼 수 있다.

표 1

특성	
환자들의 수	1,470
진단에서부터 HCT까지의 기간, 월	5.7 (0-268)
수여자 성별, 수(%) 남성/여성	833 (56.7) / 637 (43.3)
공여자 성별, 수(%) 남성/여성	977 (66.5) / 493 (33.5)
수여자 나이, 년, 중앙값 (범위)	41 (15-75)

공여자 나이, 년, 중앙값 (범위)	34 (0-70)
공여자 및 수여자의 성별 조합, 수(%) 남성에서 남성으로/여성에서 남성으로 남성에서 여성으로/여성에서 여성으로	551 (37.5) / 280 (19.0) 424 (28.8) / 213 (14.5)
질병, 수(%) AML/MDS ALL/Lymphoma/MM CML/MPN/MDS-MPN	783 (66.9) / 188 (16.1) 306 (26.2) / 56 (4.8) / 13 (1.1) 92 (7.9) / 16 (1.4) / 16 (1.4)
HCT-CI, 중앙값 (범위)	3 (0-8)
질병 위험, 수 (%) 표준 위험/고위험*	830 (56.5) / 640 (43.5)
공여자 유형, 수 (%) HLA 일치 형제 타인 HLA 반일치 가족 제대혈(Cord blood)	591 (40.2) 387 (26.4) 491 (33.4) 1 (0.1)
이식 근원(Graft source), 수 (%) 골수 말초 혈액 제대혈	472 (32.1) 997 (67.8) 1 (0.1)
컨디셔닝 강도(Conditioning intensity) 골수 소멸성(Myeloablative) 감소된 강도(Reduced-intensity)	536 (36.5) 934 (63.5)
ATG(Anti-thymocyte globulin)	903 (61.4)

- [0096] 여기서, ALL은 급성 림프모구성 백혈병(acute lymphoblastic leukemia)을 나타내고, AML은 급성 골수성 백혈병(acute myeloid leukemia)을 나타내며, CML은 만성 골수성 백혈병(chronic myeloid leukemia)을 나타내고, HCT는 조혈 세포 이식(hematopoietic cell transplantation)을 나타내며, HCT-CI는 HCT-질병 지수(HCT-comorbidity index)를 나타내고, MDS는 골수이형성 증후군(myelodysplastic syndrome)을 나타내며, MM는 다발 골수종(multiple myeloma)을 나타내고, MPN은 골수증식종양(myeloproliferative neoplasm)을 나타낼 수 있다. 표준 위험군(Standard risk group)은 첫 관해(remission)의 급성 백혈병(구제 화학요법 제외), 만성기(chronic phase)의 CML, 약물 민감성 림프종/MM, 또는 HCT에서 골수 블라스트(bone marrow blasts)가 5% 이하인 MDS를 가지는 환자로 정의될 수 있다.
- [0097] 예시적으로, 기계 학습 모델의 개발을 위하여, 미리 정의된 기간(예: 1993년 12월부터 2015년 12월까지) 동안 서울아산병원에서 혈액암으로 동종 조혈 세포 이식을 받은 1,470명의 성인(15세 이상) 환자의 데이터가 수집될 수 있다. 동종 HCT 후 장기 생존을 예측하기 위하여, 데이터는 5년 이상 생존한 환자 및 이식 후 5년 이내에 사망한 환자의 데이터를 포함할 수 있다. 데이터는, 마지막 환자의 추적 관찰 기간이 적어도 5년이 될 수 있도록, 데이터의 마감일(예: 2020년 12월)로부터 5년 이전(예: 2015년 12월 이전)에 동종 HCT를 받은 환자에 관한 데이터를 포함할 수 있다. 수여자 특성 및 공여자 특성, 질병 특징, 인간 백혈구 항원(human leukocyte antigen; HLA) 유형, 이식 정보(graft information), 투여된 컨디셔닝 약물(administered medications of conditioning), GVHD 예방 또는 지지적 치료(supportive care), 기타 실험실 데이터(other laboratory data)를 포함한 229개의 의료 피처들(예: 변수들)이 분석을 위해 수집될 수 있다.
- [0098] 2009년 12월부터 2015년 12월까지의 기간 동안, 총 1,470명의 환자들은 혈액암으로 동종 HCT를 받을 수 있다. 예시적으로, 혈액암은, 급성 골수성 백혈병, 급성 림프모구성 백혈병, 골수이형성 증후군(MDS), 만성 골수성 백혈병, 비호지킨 림프종 또는 호지킨 림프종(non-Hodgkin lymphoma and Hodgkin lymphoma), BCR-ABL1 음성 골수 증식종양(MPN), MDS 및 MPN, 및 다발성 골수종을 포함할 수 있다. 예시적으로, 데이터는 783명의 급성 골수성 백혈병 환자들, 306명의 급성 림프모구성 백혈병 환자들, 188명의 골수이형성 증후군 환자들, 92명의 만성 골수성 백혈병 환자들, 56명의 비호지킨 림프종 또는 호지킨 림프종 환자들, 16명의 BCR-ABL1 음성 MPN 환자들, 16명의 MDS 및 MPN 환자들, 및 13명의 다발성 골수종 환자들에 대한 데이터를 포함할 수 있다.
- [0099] 총 환자들의 3분의 2정도에 대응하는 995명의 환자들은 말초 혈액(peripheral blood)을 이식편(graft source)으로 받았고, 제대혈(cord blood)의 1개의 케이스가 데이터에 포함될 수 있다. 경감 강도 및 골수 소멸 컨디셔

닝(Reduced-intensity and myeloablative conditioning)은 각각 총 환자들의 63.5%인 934명의 환자들 및 총 환자들의 36.5%인 536명의 환자들에게 이용될 수 있다. ATG(Anti-thymocyte globulin)는 GVHD 예방(prophylaxis)으로서 총 환자들의 61.4%인 903명의 환자들에게 사용될 수 있다.

- [0100] 생존자의 추적 관찰 기간들의 중앙값(median)은 8.0년이고, 95% 신뢰구간(confidence interval; CI)은 7.8 이상 8.3이하의 구간일 수 있다. 생존자의 추적 관찰 기간 동안 679명의 환자들 사망했을 수 있고, 총 환자들의 추정 5년 전체 생존(overall survival; OS)은 46.2%일 수 있다. 비재발 사망(non-relapse mortality; NRM)의 2년 발생률은 17.7%, 재발의 누적 발생률(cumulative incidence of relapse; CIR)은 33.3%일 수 있다.
- [0101] 복수의 기계 학습 모델의 유형들 중 최적의 성능을 가지는 기계 학습 모델의 유형은 선택될 수 있다. 5개의 기계 학습 모델의 유형들은, 동종 조혈 세포 이식 이후 생존 또는 사망에 대한 예측의 성능에 대하여 테스트될 수 있다. 5개의 기계 학습 모델의 유형들은, 그라디언트 부스팅 머신(gradient boosting machine; GBM), 랜덤 포레스트(random forest), 딥 뉴럴 네트워크(deep neural network), 로지스틱 회귀 및 적응 부스팅(adaptive boosting; AdaBoost). 각 유형의 기계 학습 모델은 무작위로 분할된 동일한 트레이닝 데이터 세트로 테스트될 수 있다. 예를 들어, 트레이닝 데이터 세트는 전체 환자의 80%인 1,176 명의 데이터를 포함할 수 있다.
- [0102] 5개의 기계 학습 모델의 유형들 중 GBM은 0.75로 가장 높은 AUC를 가질 수 있고, 0.74의 AUC를 갖는 Random Forest, 0.65의 AUC를 갖는 Deep Neural Network, 0.70의 AUC를 갖는 Logistic Regression, 및 0.72의 AUC를 갖는 AdaBoost와 비교될 수 있다. 따라서 GBM은, 의료 피처 선택 및 최종 모델 개발을 위해 이용될 수 있다.
- [0103] GBM은 트리와 같은 여러 개의 약한 분류기들을 결합하는 앙상블 방법을 나타낼 수 있다. GBM의 목표는 경사하강법(gradient descent)을 통해 잘못 예측된 결과에 가중치를 집중하고 배치하는 것을 포함할 수 있다. GBM의 트레이닝은, 초기 트리의 데이터 세트 학습 및 잘못 예측된 레코드에 가중치 할당, 및 동일한 모델의 다음 트리의 가중치 할당된 데이터 세트 학습 및 가중치 재할당을 반복하는 것을 포함할 수 있다.
- [0104] 기계 학습 모델의 입력 데이터의 의료 피처들은 선택될 수 있다. 예를 들어, 기계 학습 모델의 최종 구조가 GBM으로 결정된 후, 기계 학습 모델의 개발에 사용된 의료 피처는 재귀적 피처 제거(Recursive Feature Elimination; RFE)를 이용하여 선택될 수 있다. 재귀적 피처 제거(RFE)는, 피처 선택(feature selection) 방법 중 하나로서, 대상 피처(target feature)를 예측하는 피처 중요도(feature importance)에 따라 각 피처에 순위를 부여하고, 기계 학습 모델에서 좋은 성능을 보이는 최소한의 지정된 수의 의료 피처들을 선택하는 것을 도울 수 있다.
- [0105] 표 2는 일 실시예에 따른 트레이닝 데이터 세트의 의료 피처들 중에서 선택된 입력 데이터의 의료 피처들을 나타낼 수 있다.

표 2

[0106]	진단 및 질병(예: AML CR1)
	질병 위험*
	진단 시 백혈구 수
	진단 시 골수의 질환(Extramedullary disease)
	동종 조혈 세포 이식 시 골수의 질환
	진단 시 핵형
	동종 조혈 세포 이식 시 핵형
	수여자의 CMV(cytomegalovirus) 혈청 상태
	공여자의 CMV(cytomegalovirus) 혈청 상태
	HCT-CI의 간(Hepatic) 스코어
	HCT-CI의 총 스코어
	전처치
	공여자 유형
	수여자 HLA 유형: A, B, C, DR, and DQ
	공여자 HLA 유형: A, B, C, DR, and DQ
	동종 조혈 세포 이식 전의 적혈구 수혈
	동종 조혈 세포 이식 전의 혈소판 수혈

- [0107] 여기서, 별(*) 표시와 관련하여, 표준 위험군(Standard risk group)은 첫 관해(remission)의 급성 백혈병(구제 화학요법 제외), 만성기(chronic phase)의 CML, 약물 민감성 림프종/MM, 또는 HCT에서 골수 블라스트(bone

marrow blasts)가 5% 이하인 MDS를 가지는 환자로 정의될 수 있다.

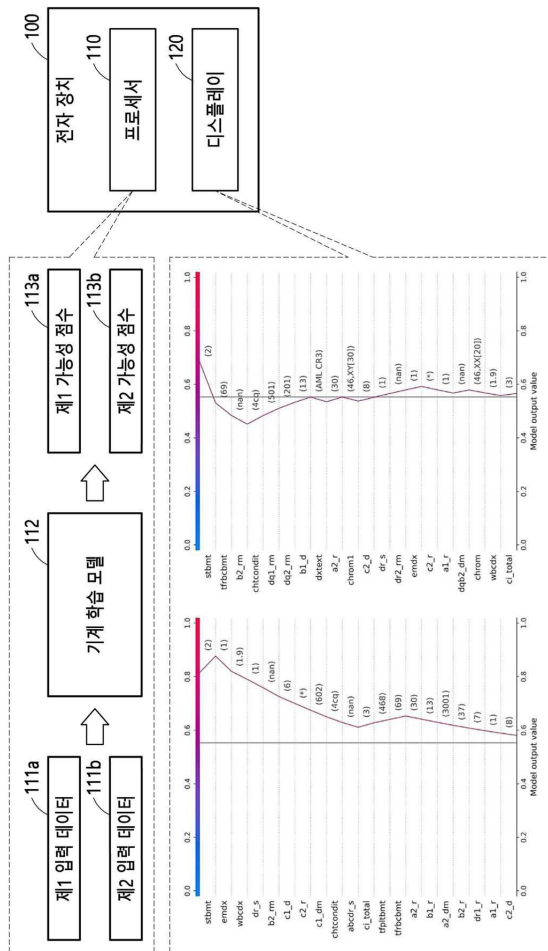
- [0108] 예를 들어, 총 229개의 의료 피처들 중에서 45개의 의료 피처들을 RFE를 이용하여 선택될 수 있다. 인간 백혈구 항원 유형(human leukocyte antigen type; HLA type)에 관하여, 공여자 및 수여자 각각의 대립유전자는 독립적인 의료 피처로 간주될 수 있다.
- [0109] 도 5는 일 실시예에 따른 기계 학습 모델의 성능을 10겹 교차 검증을 이용한 평가를 나타낸다.
- [0110] 기계 학습 모델의 예측력이 유사한 정확도로 일반화된 성능을 입증하기 위하여, 기계 학습 모델은 총 환자 코호트에서 10겹 교차 검증을 이용하여 트레이닝되고 평가될 수 있다.
- [0111] 그래프(510)는 일 실시예에 따른 기계 학습 모델의 성능에 대한 ROC 커브를 나타낸다. 예를 들어, 기계 학습 모델은 검증 코호트의 294명의 환자들로 GBM 유형의 기계 학습 모델로 구현될 수 있고, 기계 학습 모델의 입력 데이터는 선택된 의료 피처들을 포함할 수 있다. 최종 모델의 AUC 및 예측 정확도는 각각 0.788 및 0.712일 수 있다.
- [0112] 그래프(520)는 일 실시예에 따른 기계 학습 모델의 캘리브레이션 플롯(calibration plot)을 나타낸다. 예를 들어, 캘리브레이션 플롯은 검증 코호트에서 추정된 예측 위험과 관찰된 사망 위험 사이의 일치치를 보여주는 모델의 Brier 점수의 캘리브레이션 플롯을 나타낼 수 있다.
- [0113] 동종 조혈 세포 이식 후 5년 이내에 사망한 수여자가 1로 라벨링되기 때문에, 기계 학습 모델의 가능성 점수가 1에 가까울수록 사망 위험이 높을 수 있다. 가능성 점수에 기초하여 양성(positive) 또는 음성(negative)인지를 결정하기 위한 최적 임계는 수신기 작동 특성 곡선(Receiver operating characteristics curve)과 함께 Youden의 J 통계를 통해 계산될 수 있다. 예측 모델로부터, 임계는 0.5533일 수 있다. 예를 들어, 가능성 점수가 임계를 초과하면 환자는 5년 이내에 사망할 것으로 추정될 수 있다. 각 환자의 가능성 점수의 예측 확률은 총 환자들의 20%에 대응하는 무작위로 선택된 환자 코호트로 테스트되어 생성된 모델들 10개 중 1개의 모델을 선택함으로써 가능한 편향을 줄일 수 있다.
- [0114] 도 6은 일 실시예에 따른 기계 학습 모델의 가능성 점수에 기초하여 분할된 환자 클래스에 따른 전체 생존, 재발의 발생률, 및 비재발 사망률을 나타낸다.
- [0115] 데이터 세트의 환자들은 기계 학습 모델에 의하여 출력된 가능성 점수에 기초하여 균등하게 3개의 환자 클래스들로 분할될 수 있다. 데이터 세트의 환자들은, 저위험군, 중간위험군, 및 고위험군 중 하나의 환자 클래스로 분류될 수 있다.
- [0116] 참고로, 범주형 변수는 카이제곱 검정 또는 Fisher의 정확 검정을 통해 비교될 수 있고, 연속형 변수는 Mann-Whitney U-검정 또는 Student's t-검정을 통해 비교될 수 있다. 전체 생존(OS)은 Kaplan-Meier 방법에 의해 계산될 수 있고, 결과 생존 곡선(resulting survival curve)은 로그 순위 테스트(log-rank test)를 통해 비교될 수 있다. NRM과 CIR은 경쟁 위험(competing risks)에 대한 누적 발생 함수(cumulative incidence function)를 이용하여 평가될 수 있고, Gray의 방법과 R 버전 3.6.3(R Foundation for Statistical Computing, Vienna, Austria)을 통해 비교될 수 있다. 통계는 SPSS 버전 24(IBM Corporation, Somers, NY, USA)를 이용하여 산출될 수 있고, 그래프는 GraphPad Prism 버전 9.1.2(GraphPad Software, Inc., La Jolla, CA, USA)를 이용하여 생성될 수 있다. 모든 분석에서, P 밸류는 양측(two-tailed)일 수 있고, 0.05 미만의 P 밸류는 통계적으로 유의한 것으로 간주될 수 있다.
- [0117] 그래프(610)는 각 환자 클래스에 대한 동종 조혈 세포 이식 이후 기간에 따른 전체 생존을 나타낼 수 있다. 추정된 5년 전체 생존(5-year Overall Survival)는 저위험군(도 6에서 low로 표시됨)에서 70.3%, 중간위험군(도 6에서 intermediate로 표시됨)에서 42.6%, 고위험군(도 6에서 high로 표시됨)에서 14.9%일 수 있다. 3개의 환자 클래스들 간의 5년 전체 생존에 대한 P 밸류는 0.001 미만일 수 있다.
- [0118] 비재발 사망률(NRM) 및 재발의 발생률이 가능성 점수에 기초하여 예측될 수 있는지 평가하기 위하여, 각 환자 클래스에 대한 비재발 사망률(NRM) 및 재발의 발생률은 분석될 수 있다. 그래프(620) 및 그래프(630)에 나타난 바와 같이, 높은 가능성 점수는 높은 재발의 발생률(CIR) 및 비재발 사망률(NRM) 모두와 유의하게 연관될 수 있다. 3개의 환자 클래스들 간의 재발의 발생률에 대한 P 밸류는 0.001 미만일 수 있다. 3개의 환자 클래스들 간의 비재발 사망률에 대한 P 밸류는 0.016일 수 있다. 추정된 2년 재발의 발생률(CIR)은 저위험군에서 11.3%, 중간위험군에서 22.4%, 고위험군에서 53.1%일 수 있다. 2년 비재발 사망률(NRM)은 저위험군에서 9.3%, 중간 위험군에서 17.3%, 고위험군에서 각각 25.0%일 수 있다.

- [0119] 혈액암에서 동종 조혈 세포 이식 후 장기 생존은 여러 요인에 의해 영향을 받을 수 있지만 주로 질병 재발 및 비재발 사망률(NRM)에 의존할 수 있다. 질병 상태(disease status), 유전적 위험(genetic risk), 전처치(conditioning regimen), 환자의 동반 질환(comorbidities of patients), HLA 매칭의 정도(degree of HLA matching), 및 환자 및 기증자 연령(patient and donor age)을 포함한 여러가지 의료 피처들은 질병 재발, GVHD, 이식 또는 치료 관련 독성과 연관될 수 있고, 이러한 결과들은 이식 후 생존과 밀접하고 상호 관련될 수 있다. 그러나, 종래의 통계적 방법은 수여자 및 공여자의 HLA 대립유전자 간의 관계 등 의료 피처 간의 상호작용 또는 각 의료 피처의 구체적인 값에 따른 차이를 고려한 분석에 적합하지 않을 수 있다. 일 실시예에 따른 기계 학습 모델은 동종 조혈 세포 이식 후 결과를 예측하고, 적절한 수여자, 기증자, 및 이식의 계획 중 적어도 하나를 선택하는 지침을 제시하는 효과적인 대안이 될 수 있다.
- [0120] 동종 조혈 세포 이식 후 장기 생존을 위한 기계 학습 모델의 유형(예: GBM) 및 선택된 의료 피처들을 이용하여 기계 학습 모델 및 가능성 점수는 개발될 수 있다. 일 실시예에 따른 기계 학습 모델은 이식 후 결과를 추정하는 종래의 비교 실시예에 따른 기계 학습 모델과 비교하여 더 나은 성능을 보여주는 0.788의 AUC를 가질 수 있다. 비교 실시예 1에 따른 기계 학습 모델은 0.701의 AUC를 가질 수 있고, 0.646의 EBMT 점수보다 유의하게 우수한 동종 조혈 세포 이식 후 100일 사망률을 예측하는 AL-EBMT 모델을 나타낼 수 있다. 일본 조혈 세포 이식 학회에서 개발된 급성 GVHD의 기계 학습 모델은 0.62의 AUC를 가질 수 있다. Fuse et al에 따른 기계 학습 모델은 동종 조혈 세포 이식 후 1년 이내에 재발을 예측하는 모델로서, 0.667의 AUC를 가질 수 있다. 대부분의 모델들은 교대 결정 트리 알고리즘(alternating decision tree algorithm)을 적용하여 개발되고, 의료 피처들(예: 변수)은 연구자에 의하여 선택될 수 있다. 일 실시예에 따른 기계 학습 모델은, 연구자의 의견 또는 기존의 통계적 분석에 기초하여 미리 선택된 의료 피처를 사용하는 것이 아니라, 기계 학습 모델에 대한 피처 중요도에 기초하여 선택된 의료 피처들을 이용하여 개발될 수 있다. 재귀적 피처 제거(RFE)를 통해, 기계 학습 모델의 예측 성능이 저하되지 않는 최소 요구 의료 피처의 개수는 추출될 수 있다. 일 실시예에 따른 기계 학습 모델은 재귀적 피처 제거를 통해 선택된 의료 피처들을 이용하는 점에서, 비교 실시예들에 따른 기계 학습 모델이 임상적으로 선택된 의료 피처를 이용하는 것과 중요한 차이점을 가질 수 있다. 일 실시예에 따른 프로세서는 가능한 모든 의료 피처들(예: 수집된 의료 피처들 전부)을 이용하여 초기의 임시 기계 학습 모델을 구축하고, 가능성 점수에 거의 영향을 미치지 않는 의료 피처를 점진적으로 제거할 수 있다. 프로세서는, 의료 피처의 선택에 대한 바이어스를 감소시킬 수 있고, 의료 피처들 간의 가능한 상관관계를 반영함으로써 기계 학습 모델이 더 높은 AUC를 가질 수 있도록 기여할 수 있다. 흥미롭게도, 일 실시예에 따른 기계 학습 모델을 위해 선택된 의료 피처는 수여자 및 공여자의 각 HLA 대립 유전자 유형을 포함할 수 있다. 미스매치 정도, 미스매치 방향, 대립유전자 유형의 계산보다 수여자 및 공여자의 각 HLA 대립유전자에 대한 원시 값(raw value)은 사용될 수 있기 때문에, 가능성 점수에 영향을 미치는 대립유전자 간의 상호작용은 통합될 수 있다.
- [0121] 동종 조혈 세포 이식을 계획 중인 수여자를 위하여 기계 학습 모델을 이용하려면, 다양한 동종 조혈 세포 이식 케이스들에 따른 예상 결과들을 비교할 수 있는 도구가 필요할 수 있다. 예측 점수는 동종 조혈 세포 이식 후 관찰된 생존 및 추정된 생존이 잘 일치하는 생존 확률을 정량화할 수 있다. 피처 기여도(예: 새플리 값)는 각 의료 피처의 영향력의 시각화를 가능하게 할 수 있다. 이식 관련 자원(예: 공여자, 수여자, 및 이식 계획)의 우선순위는 피처 기여도에 기초하여 지정될 수 있다. 일 실시예에 따른 전자 장치는, 동종 조혈 세포 이식을 계획을 세우는 의료인에 의하여 이용되는 경우, 보다 더 높은 생존 가능성을 달성할 것으로 예상되는 공여자 또는 전처치과 같은 요인들을 선택할 수 있도록 각 요인의 영향력을 정량화하고 시각화할 수 있다.
- [0122] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

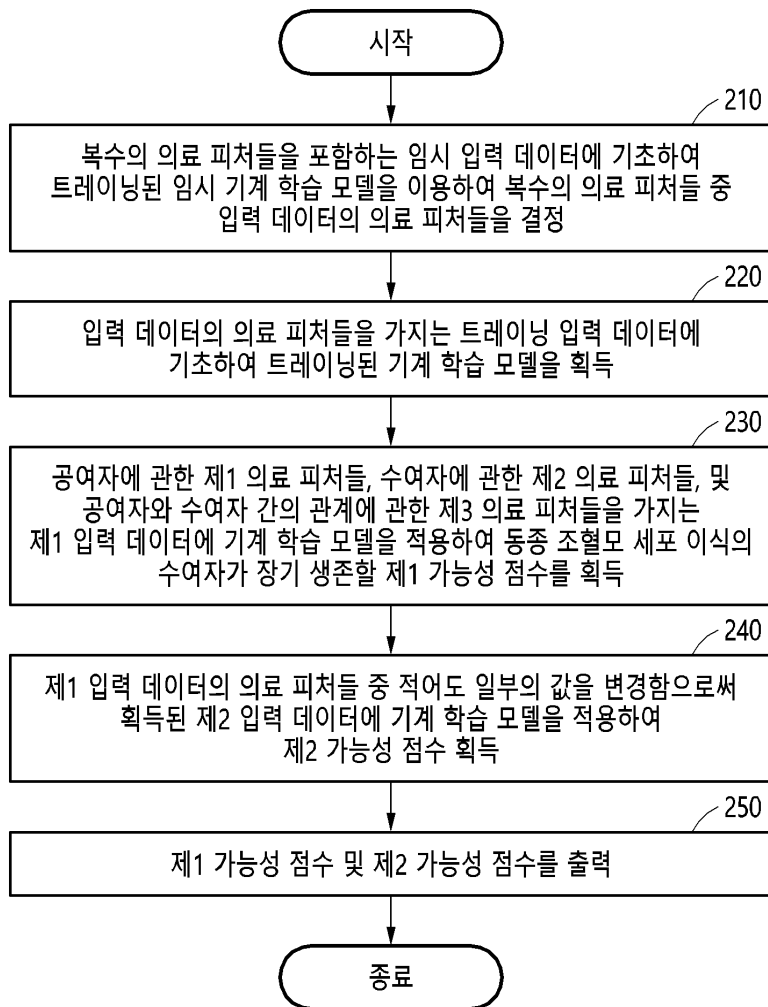
- [0123] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.
- [0124] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있으며 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [0125] 위에서 설명한 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 또는 복수의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0126] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 이를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.
- [0127] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

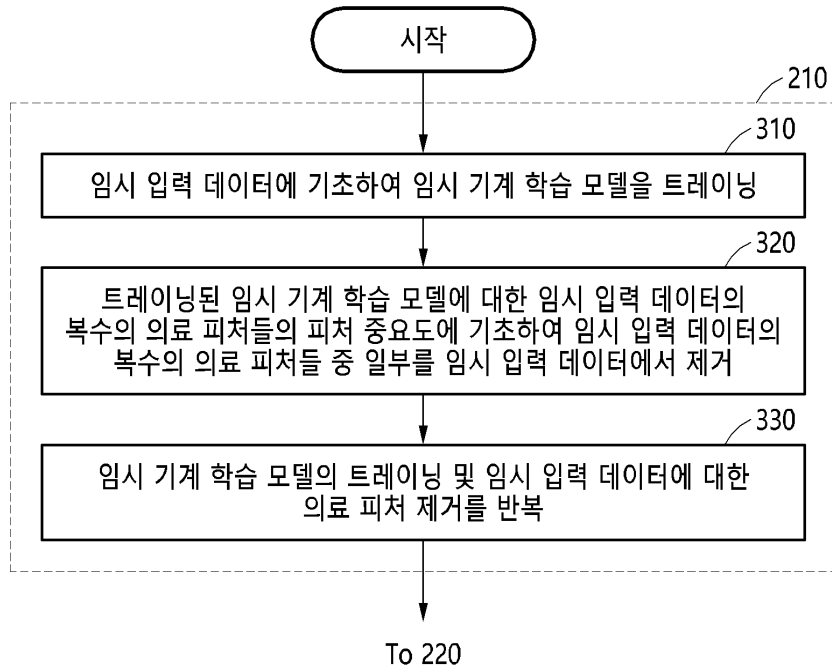
도면1



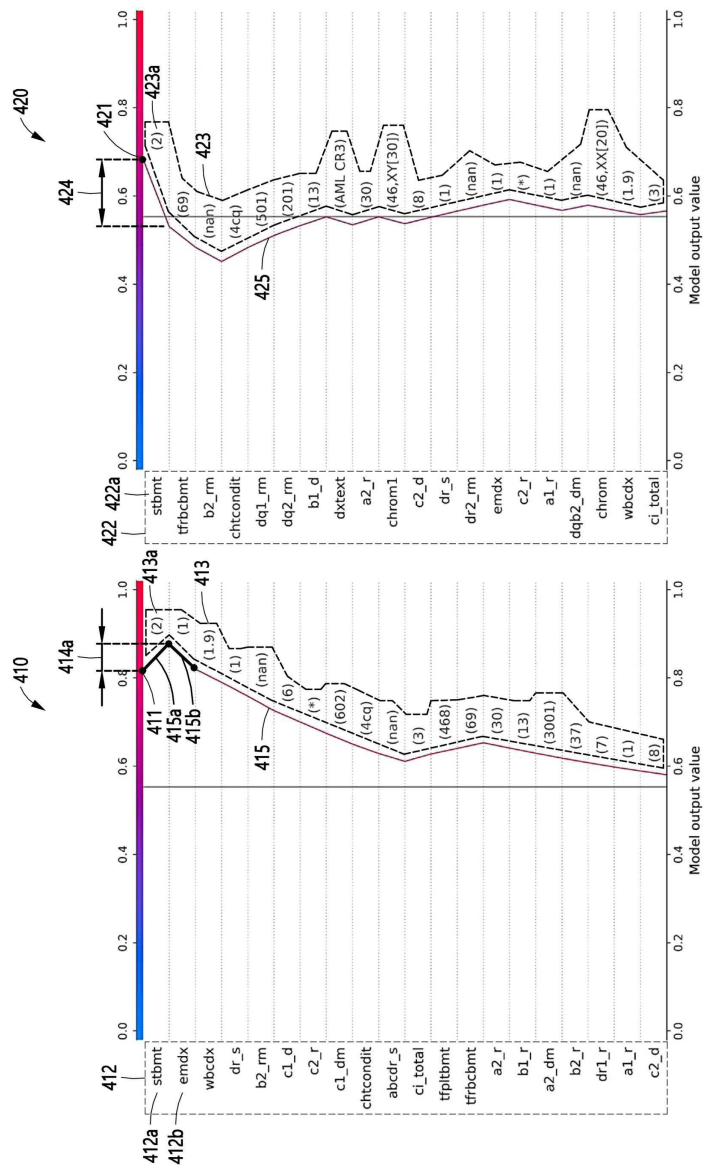
도면2



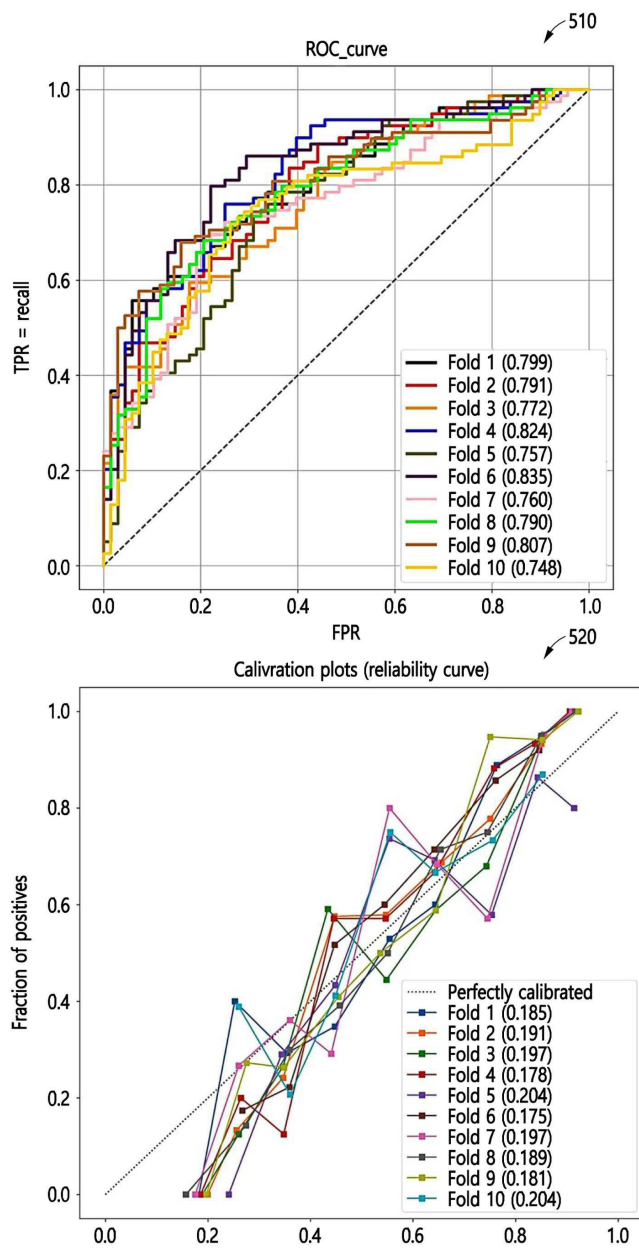
도면3



도면4



도면5



도면6

