



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0042278
(43) 공개일자 2024년04월02일

(51) 국제특허분류(Int. Cl.)

G16H 20/10 (2018.01) G06N 20/00 (2019.01)
G16H 10/00 (2018.01) G16H 50/70 (2018.01)
G16H 70/00 (2018.01)

(52) CPC특허분류

G16H 20/10 (2021.08)
G06N 20/00 (2021.08)

(21) 출원번호 10-2022-0120539

(22) 출원일자 2022년09월23일

심사청구일자 2022년09월23일

(71) 출원인

재단법인 아산사회복지재단

서울특별시 송파구 올림픽로43길 88 (풍납동)

울산대학교 산학협력단

울산광역시 남구 대학로 93(무거동)

충남대학교병원

대전광역시 중구 문화로 282 (대사동)

(72) 발명자

김영학

서울특별시 강남구 선릉로 120, 3동 1102호 (대치동, 개포1,2차우성아파트)

전태준

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

(뒷면에 계속)

(74) 대리인

특허법인 무한

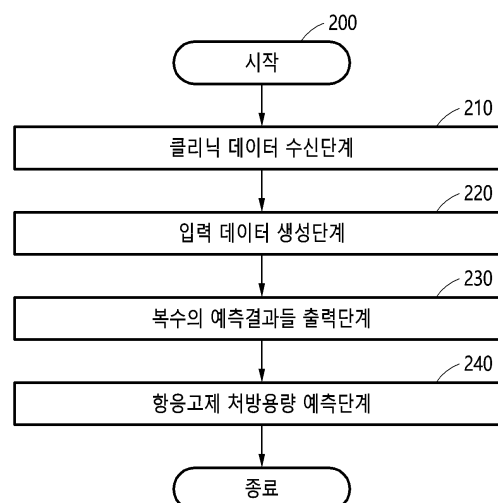
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 머신러닝 기술을 활용한 항응고제 초기 적정 용량 예측 방법 및 장치

(57) 요약

일 실시예에 따른 프로세서에 의해 구현되는 항응고제 처방용량을 예측하는 방법은 환자의 클리닉 데이터를 수신하는 단계; 상기 클리닉 데이터에 기초하여 입력 데이터를 생성하는 단계; 상기 항응고제의 처방 예상 범위가 단위 용량으로 분할된 복수의 처방용량 구간들의 각각에 대응하는 서브 기계학습 모델을 포함하는 제1 기계학습 모델에 상기 입력 데이터를 적용함으로써, 상기 복수의 처방용량 구간들에 대한 복수의 예측 결과들을 출력하는 단계; 및 상기 출력된 복수의 예측 결과들에 기초하여 상기 환자의 항응고제 처방용량을 예측하는 단계를 포함할 수 있다.

대표도 - 도2



(52) CPC특허분류

G16H 10/00 (2021.08)

G16H 50/70 (2018.01)

G16H 70/00 (2021.08)

(72) 발명자

최희정

서울특별시 강남구 봉은사로49길 21-10, 202호(논현동, 블루하우스(Blue house))

노재형

세종특별자치시 보람로 15, 911동 301호(도담동, 도람마을9단지)

강희준

경기도 성남시 중원구 둔촌대로 300, 1004동 1004호(하대원동, 아튼빌아파트)

권한솔

서울특별시 노원구 초안산로5길 22, 105동 202호(월계동, 대우아파트)

안임진

경기도 의정부시 진등로19번길 20, 901동 1207호(녹양동, 녹양동 신도9차 전원마을아파트)

김윤하

서울특별시 성동구 서울숲길 25, 102동 304호(성수동1가, 현대아파트)

서혜람

서울특별시 강동구 양재대로124길 67-11, B102호(길동, 금석빌라)

조하나

서울특별시 송파구 오금로11길 41, 704호(방이동)

김민경

경기도 의정부시 오목로 36, 103동 1502호(용현동, 현대아파트)

한지예

서울특별시 양천구 목동동로 10, 1110동 601호(신정동, 목동신시가리아파트11단지)

기가은

서울특별시 서초구 서초대로16길 22, 302호(방배동, 삼풍맨션)

박서현

서울특별시 서초구 서초대로33길 84-24, B동 402호(방배동, 방배4차상지리츠빌)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711138158

과제번호 KMDF_PR_20200901_0097-01

부처명 다부처

과제관리(전문)기관명 (재단)범부처전주기의료기기연구개발사업단

연구사업명 범부처전주기의료기기연구개발사업(R&D)(과기정통부, 복지부, 산업부)

연구과제명 (주관)EMR기반 환자중심 진단치료 지원용 초지능 의료기기 소프트웨어 개발 및 실증

기 여 율 1/2

과제수행기관명 서울아산병원

연구기간 2021.03.01 ~ 2022.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호 1465035447

과제번호 HR21C0198010022

부처명 보건복지부

과제관리(전문)기관명 한국보건산업진흥원

연구사업명 연구중심병원 육성 R&D

연구과제명 의료현장 통합 지식베이스 및 인공지능형 가상 임상시험 시스템 구축

기 여 율 1/2

과제수행기관명 서울아산병원

연구기간 2022.01.01 ~ 2022.12.31

명세서

청구범위

청구항 1

프로세서에 의해 구현되는 항응고제 처방용량을 예측하는 방법에 있어서,
환자의 클리닉 데이터를 수신하는 단계;
상기 클리닉 데이터에 기초하여 입력 데이터를 생성하는 단계;
상기 항응고제의 처방 예상 범위가 단위 용량 구간으로 분할된 복수의 처방용량 구간들의 각각에 대응하는 서브 기계학습 모델들을 포함하는 제1 기계학습 모델에 상기 입력 데이터를 적용함으로써, 상기 복수의 처방용량 구간들에 대한 복수의 예측 결과들을 출력하는 단계; 및
상기 출력된 복수의 예측 결과들에 기초하여 상기 환자의 항응고제 처방용량을 예측하는 단계
를 포함하는
항응고제 처방용량 예측 방법.

청구항 2

제1항에 있어서,
상기 입력 데이터를 생성하는 단계는,
상기 클리닉 데이터에 기초하여 인구통계학적 정보(demographic information), 활력징후(Vital sign), 진단정보 및 복용정보를 포함하는 상기 입력 데이터로 생성하는 단계
를 포함하는 항응고제 처방용량 예측 방법.

청구항 3

제1항에 있어서,
복수의 처방용량 구간들은 상기 단위 용량 구간들에 기초하여 등간격으로 구분되는
항응고제 처방용량 예측 방법.

청구항 4

제1항에 있어서,
상기 복수의 예측 결과들을 출력하는 단계는,
상기 서브 기계학습 모델들의 각각에서 해당하는 처방용량 구간에서의 처방 여부를 지시하는 출력을 상기 복수의 예측 결과들로서 생성하는 단계
를 포함하고,
상기 처방용량을 예측하는 단계는,
처방용량 구간 별로 생성된 처방 여부를 지시하는 출력들에 기초하여 상기 처방용량을 결정하는 단계

를 포함하는 항응고제 처방용량 예측 방법.

청구항 5

제4항에 있어서,

상기 처방용량을 결정하는 단계는,

상기 생성된 출력 중 해당하는 처방용량 구간에서 처방하는 것으로 결정된 출력의 개수에 대응하는 용량을 상기 처방용량으로 결정하는 단계

를 포함하는 항응고제 처방용량 예측 방법.

청구항 6

제1항에 있어서,

클리닉 데이터 중으로부터 추출된 일부 데이터를 포함하는 트레이닝 입력 및 참값(ground truth) 용량을 라벨로 가지는 트레이닝 출력의 쌍을 포함하는 트레이닝 데이터를 이용하여 상기 제1 기계학습 모델을 트레이닝시키는 단계

를 더 포함하는 항응고제 처방용량 예측 방법.

청구항 7

제6항에 있어서,

상기 트레이닝 데이터는,

트레이닝을 위한 클리닉 데이터 세트 중 해당하는 환자의 퇴원 시 의사에 의해 처방된 항응고제 처방용량이 참값용량으로서 라벨링된 트레이닝 출력을 포함하는

항응고제 처방용량 예측 방법.

청구항 8

제7항 있어서,

상기 라벨링된 트레이닝 출력의 참값용량과 상기 제1 기계학습 모델에서 산출한 항응고제 처방용량을 비교하여 로스를 결정하는 단계;

상기 로스가 임계 범위를 초과하는 것에 기초하여 가중치를 조정하는 단계; 및

상기 로스가 임계 범위 내로 될 때까지 로스의 결정과 가중치의 조정을 반복하는 단계

를 더 포함하는 항응고제 처방용량 예측 방법.

청구항 9

제1항에 있어서,

상기 제1 기계학습 모델과 다른 구조를 가지는 제2 기계학습 모델을 트레이닝 데이터를 이용하여 트레이닝시키는 단계

를 더 포함하고,

상기 처방용량을 예측하는 단계는,

상기 제2 기계학습 모델에 기초하여, 상기 입력 데이터로부터 항응고제의 처방용량을 추정하는 단계; 및

상기 제1 기계학습 모델을 이용하여 결정된 항응고제의 처방용량과 상기 제2 기계학습 모델을 이용하여 추정된 항응고제의 처방용량에 기초하여 최종 항응고제의 처방용량을 결정하는 단계

를 포함하는 항응고제 처방용량 예측 방법.

청구항 10

제1항 내지 제9항 중 어느 한 항의 방법을 수행하기 위한 명령어를 포함하는 하나 이상의 컴퓨터 프로그램을 저장한 컴퓨터 판독 가능 기록 매체.

청구항 11

전자 장치에 있어서,

항응고제의 처방 예상 범위가 단위 용량으로 분할된 복수의 처방용량 구간들의 각각에 대응하는 서브 기계학습 모델들을 포함하는 제1 기계학습 모델을 저장한 메모리; 및

환자의 클리닉 데이터에 기초하여 입력 데이터를 생성하고, 상기 제1 기계학습 모델에 상기 입력 데이터를 적용함으로써 상기 복수의 처방용량 구간들에 대한 복수의 예측 결과들을 출력하며, 상기 출력된 복수의 예측 결과들에 기초하여 상기 환자의 항응고제 처방용량을 예측하는 프로세서

를 포함하는 전자 장치.

청구항 12

제11항 있어서,

상기 입력 데이터는 상기 클리닉 데이터에 기초하여 인구통계학적정보, 활력징후, 진단정보 및 복약정보를 포함하는 전자 장치.

청구항 13

제11항 있어서,

상기 프로세서는,

상기 서브 기계학습 모델들의 각각에서 해당하는 처방용량 구간에서의 처방 여부를 지시하는 출력을 생성하고,

상기 생성된 출력 중 해당하는 처방용량 구간에서 처방하는 것으로 결정된 출력의 개수에 대응하여 처방용량을 예측하는 전자 장치.

청구항 14

제11항 있어서,

상기 제1 기계학습 모델은,

클리닉 데이터 중으로부터 추출된 일부 데이터를 포함하는 트레이닝 입력 및 환자의 퇴원 시 의사에 의해 처방된 항응고제 처방용량인 참값 용량을 라벨로 가지는 트레이닝 출력의 쌍을 포함하는 트레이닝 데이터를 이용하여 트레이닝되는 전자 장치.

청구항 15

제14항 있어서,

상기 프로세서는

상기 라벨링된 트레이닝 출력의 참값용량과 상기 제1 기계학습 모델에서 산출한 항응고제 처방용량을 비교하여 로스를 결정하고,

상기 로스가 임계범위를 초과하는 것에 기초하여 가중치를 조정하며, 상기 로스가 임계 범위 내로 될 때까지 로스의 결정과 가중치의 조정을 반복하는

전자 장치.

청구항 16

제11항 있어서,

상기 메모리는 제2 기계학습 모델을 더 저장하고,

상기 프로세서는 상기 제2 기계학습 모델을 트레이닝 데이터로 트레이닝하고, 상기 제1 기계학습 모델에서 결정된 항응고제의 처방용량과 상기 제2 기계학습 모델에서 결정된 항응고제의 처방용량에 기초하여 최종 항응고제의 처방용량을 결정하는 전자 장치.

발명의 설명

기술 분야

[0001] 항응고제 처방용량 예측 방법 및 장치에 연관되며, 보다 특정하게는 항응고제의 처방 예상 범위가 단위 용량 구간으로 분할된 복수의 처방용량 구간들의 각각에 대응하는 서브 기계학습 모델들의 출력에 기초하여 항응고제 처방용량을 예측하는 방법 및 장치에 연관된다.

배경 기술

[0003] 항응고제 요법의 목적은 자연 출혈은 피하면서 혈전은 방지하도록 혈액의 응고 능력을 감소시키는 것이다. 경구용 항응고제의 일종인 와파린은 예를 들어 정맥 혈전 및 그의 확대, 및 폐색전증의 예방 및/또는 치료를 위해, 심방세동 및/또는 심장판막 대치술과 관련된 혈전색전성 합병증의 예방 및/또는 치료를 위해, 또한 사망, 심근경색 재발, 및 예를 들어 심근경색 후의 졸증 또는 전신성 색전형성과 같은 혈전색전성 사고의 위험을 감소시키기 위해 흔히 처방될 수 있다.

[0004] 그러나 와파린은 좁은 치료지수(narrow therapeutic index)를 가지고 있다. 치료지수란, 약물용량에 대한 독성을 나타내는 약물용량의 비를 의미한다. 이에 따라 좁은 치료지수는 의약품 복용량 조절을 타이트하게 조정해야 한다. 그러나 치료지수는 개인간 편차가 커 복용량을 결정하는데 어려움이 있다. 만일 부적절한 용량의 와파린을 복용하는 경우, 합병증 발생 또는 심각한 출혈부작용의 위험이 증가된다.

[0005] 환자마다 적절한 와파린 용량을 결정하기 위해, 병원에서는 2~4주에 한 번씩 주기적인 혈액검사를 하고, 적절한 용량을 찾은 후에는 3~6개월에 한 번씩 혈액검사를 진행할 수 있다. 혈액검사에서는 혈액응고 측정의 표시로서 와파린 용량을 조절하기 위한 INR(international normalized ratio) 수치를 확인할 수 있다. 정상인의 INR 수치는 1이며 폰탄(Fontan)수술 후에는 이를 2~2.7 사이로 유지하는 것이 바람직할 수 있다. INR 수치가 기준 수치보다 더 높으면 응고기능이 너무 많이 억제된 것을 의미하며 이럴 경우 출혈의 위험이 있으므로 와파린의 용량을 줄여야 한다. 반대로 INR 수치가 기준 수치보다 낮으면 응고억제 효과가 부족하므로 용량을 올려야 한다.

[0006] 특히 와파린은 간에서 와파린 대사를 책임지는 두 가지 유전자의 변이에 따라 적절한 용량이 다른 것으로 볼 수

있다. CYP2C9 유전자에 특정 변이가 생기면, 와파린 대사가 감소되고 출혈 부작용이 더 크기 때문에 상대적으로 적은 와파린 용량을 복용해야 한다. VKORC1 유전자에 특정 변이가 생기면, 와파린 감수성이 증가하기 때문에 와파린 용량을 감량하여야 한다. 이러한 결과가 밝혀짐으로써, 2010년 FDA는 CYP2C9와 VKORC1의 유전형 검사가 와파린 적정 용량 결정에 유용할 것으로 판단하여 CYP2C9와 VKORC1 유전형 조합에 따른 와파린 용량 가이드라인을 만들었다. 또한, 와파린 치료 시 개인별 적정 용량의 예측을 위해 CYP2C9 및 VKORC1 유전형검사(genotyping)를 시행하는 것이 권고될 수 있다.

[0007] 따라서, 유전형검사 데이터와 임상데이터를 모두 사용하여 개인별 와파린 적정 용량을 예측하는 머신 러닝(machine learning) 및 딥 러닝 알고리즘(deep learning algorithm)이 개발되어 왔다. [use of pharmacogenetic and clinical factors to predict the maintenance dose of warfarin, Brian F Gage, Charles Eby, Paul E Milligan, Gerald A Banet, Jill R Duncan, Howard L McLeod, 2004 Jan;91(1):87-94] 논문에서 개발한 약물 유전학적 알고리즘(pharmacogenetic algorithm)은 와파린 적정 용량을 예측함으로써, 와파린 과다복용 환자를 6.5% 감소시킬 수 있다는 것을 확인할 수 있다. [Artificial neural network-based pharmacogenomic algorithm for warfarin dose optimization, Addepalli Pavani, Shaik Mohammad Naushad, Rajasekar Manoj Kumar, Murali Srinath, Amaresh Rao Malempati, Vijay Kumar Kutala, 2016;17(2):121-31] 논문에서 개발한 인공 신경망(artificial neural network) 모델은 와파린 용량 변동성을 93.5% 설명할 수 있다는 것을 확인할 수 있다. 그 외에도 다양한 머신러닝(ML) 모델들이 와파린의 변동성을 43~54% 설명할 수 있음을 확인할 수 있다.

발명의 내용

해결하려는 과제

[0009] 일 실시예에 따른 항응고제 처방용량을 예측하는 방법 및 장치는 항응고제의 처방용량을 예측할 수 있다.

[0010] 다만, 기술적 과제는 상술한 기술적 과제들로 한정되는 것은 아니며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0012] 일 실시예에 따른 프로세서에 의해 구현되는 항응고제 처방용량을 예측하는 방법은, 환자의 클리닉 데이터를 수신하는 단계; 상기 클리닉 데이터에 기초하여 입력 데이터를 생성하는 단계; 상기 항응고제의 처방 예상 범위가 단위 용량으로 분할된 복수의 처방용량 구간들의 각각에 대응하는 서브 기계학습 모델들을 포함하는 제1 기계학습 모델에 상기 입력 데이터를 적용함으로써, 상기 복수의 처방용량 구간들에 대한 복수의 예측 결과들을 출력하는 단계; 및 상기 출력된 복수의 예측 결과들에 기초하여 상기 환자의 항응고제 처방용량을 예측하는 단계를 포함할 수 있다.

[0013] 상기 입력 데이터를 생성하는 단계는, 상기 클리닉 데이터에 기초하여 인구통계학적 정보(demographic information), 활력징후(Vital sign), 진단정보 및 복용정보를 포함하는 상기 입력 데이터로 생성하는 단계를 포함할 수 있다.

[0014] 상기 단위 용량 구간들은 등간격으로 구분될 수 있다.

[0015] 상기 복수의 예측 결과들을 출력하는 단계는, 상기 서브 기계학습 모델들의 각각에서 해당하는 처방용량 구간에서의 처방 여부를 지시하는 출력을 생성하는 단계를 포함하고, 상기 처방용량을 예측하는 단계는, 상기 생성된 출력 중 해당하는 처방용량 구간에서 처방하는 것으로 결정된 출력의 개수에 대응하여 처방용량을 예측하는 단계를 포함할 수 있다.

[0016] 상기 제1 기계학습 모델은, 클리닉 데이터 중으로부터 추출된 일부 데이터를 포함하는 트레이닝 입력 및 참값(ground truth) 처방용량을 라벨로 가지는 트레이닝 출력의 쌍을 포함하는 트레이닝 데이터를 이용하여 트레이닝될 수 있다.

[0017] 상기 트레이닝 데이터를 제2 기계학습 모델에 추가로 트레이닝시켜, 상기 제1 기계학습 모델에서 결정된 항응고제의 처방용량과 상기 제2 기계학습 모델에서 결정된 항응고제의 처방용량에 기초하여 최종 항응고제의 처방용량을 결정할 수 있다.

- [0018] 상기 트레이닝 데이터는, 트레이닝을 위한 클리닉 데이터 세트 중 해당하는 환자의 퇴원 시 의사에 의해 처방된 항응고제 처방용량이 참값용량으로서 라벨링된 트레이닝 출력을 포함할 수 있다.
- [0019] 상기 라벨링된 트레이닝 출력의 참값용량과 상기 제1 기계학습 모델 및 제2 기계학습 모델에서 산출한 항응고제 처방용량을 각각 비교하여 로스를 결정하는 단계; 상기 로스가 임계 범위를 초과하는 것에 기초하여 가중치를 조정하는 단계; 및 상기 로스가 임계 범위 내로 될 때까지 로스의 결정과 가중치의 조정을 반복하는 단계를 더 포함할 수 있다.
- [0020] 상기 제1 기계학습 모델과 다른 구조를 가지는 제2 기계학습 모델을 트레이닝 데이터를 이용하여 트레이닝 시키는 단계를 더 포함하고, 상기 처방용량을 예측하는 단계는, 상기 제2 기계학습 모델에 기초하여, 상기 입력 데이터로부터 항응고제의 처방용량을 추정하는 단계; 및 상기 제1 기계학습 모델을 이용하여 결정된 항응고제의 처방용량과 상기 제2 기계학습 모델을 이용하여 추정된 항응고제의 처방용량에 기초하여 최종 항응고제의 처방용량을 결정하는 단계를 포함할 수 있다.
- [0021] 일 실시예에 따른 하나 이상의 컴퓨터 프로그램을 저장한 컴퓨터 판독 가능 기록 매체에 있어서, 방법을 수행하기 위한 명령어를 포함할 수 있다.
- [0022] 일 실시예에 따른 전자 장치는, 항응고제의 처방 예상 범위가 단위 용량으로 분할된 복수의 처방용량 구간들의 각각에 대응하는 서브 기계학습 모델들을 포함하는 제1 기계학습 모델을 저장한 메모리; 및 환자의 클리닉 데이터에 기초하여 입력 데이터를 생성하고, 상기 제1 기계학습 모델에 상기 입력 데이터를 적용함으로써 상기 복수의 처방용량 구간들에 대한 복수의 예측 결과들을 출력하며, 상기 출력된 복수의 예측 결과들에 기초하여 상기 환자의 항응고제 처방용량을 예측하는 프로세서를 포함할 수 있다.
- [0023] 상기 입력 데이터는 상기 클리닉 데이터에 기초하여 인구통계학적정보, 활력징후, 진단정보 및 복약정보를 포함할 수 있다.
- [0024] 상기 프로세서는, 상기 서브 기계학습 모델들의 각각에서 해당하는 처방용량 구간에서의 처방 여부를 지시하는 출력을 생성하고, 상기 생성된 출력 중 해당하는 처방용량 구간에서 처방하는 것으로 결정된 출력의 개수에 대응하여 처방용량을 예측할 수 있다.
- [0025] 상기 제1 기계학습 모델은, 클리닉 데이터 중으로부터 추출된 일부 데이터를 포함하는 트레이닝 입력 및 환자의 퇴원 시 의사에 의해 처방된 항응고제 처방용량인 참값 용량을 라벨로 가지는 트레이닝 출력의 쌍을 포함하는 트레이닝 데이터를 이용하여 트레이닝될 수 있다.
- [0026] 상기 프로세서는 상기 라벨링된 트레이닝 출력의 참값용량과 상기 제1 기계학습 모델에서 산출한 항응고제 처방용량을 비교하여 로스를 결정하고, 상기 로스가 임계범위를 초과하는 것에 기초하여 가중치를 조정하며, 상기 로스가 임계 범위 내로 될 때까지 로스의 결정과 가중치의 조정을 반복할 수 있다.
- [0027] 상기 메모리는 제2 기계학습 모델을 더 저장하고, 상기 프로세서는 상기 제2 기계학습 모델을 트레이닝 데이터로 트레이닝하고, 상기 제1 기계학습 모델에서 결정된 항응고제의 처방용량과 상기 제2 기계학습 모델에서 결정된 항응고제의 처방용량에 기초하여 최종 항응고제의 처방용량을 결정할 수 있다.

발명의 효과

- [0029] 일 실시예에 따른 항응고제 처방용량을 예측하는 방법 및 장치는 항응고제의 처방용량을 예측함에 있어 비용 및 시간을 효율적으로 사용할 수 있다.

도면의 간단한 설명

- [0031] 도1은 일 실시예에 따른 항응고제 처방용량 예측 장치의 예시적인 구조를 도시한다.
- 도2는 일 실시예에 따른 항응고제 처방용량 예측 방법의 흐름도이다.
- 도3은 일 실시예에 따른 클리닉 데이터 전체항목을 SHAP 기법을 이용하여 특징 변수를 시각화한 예시도이다.
- 도4는 일 실시예에 따른 제1 기계학습 모델의 구조에서 복수의 예측 결과들이 출력되는 흐름을 나타내는 모식도

이다.

도5는 일 실시예에 따른 제1 기계학습 모델의 구조에서 항응고제 처방용량을 예측하는 예시도이다.

도6은 일 실시예에 따른 제1 기계학습 모델의 구조에서 항응고제 처방용량을 예측하는 또 다른 예시도이다.

도7은 일 실시예에 따른 제2 기계학습 모델의 구조에서 항응고제 처방용량을 예측하는 흐름을 나타내는 모식도이다.

도8은 일 실시예에 따른 제1 기계학습 모델의 구조와 제2 기계학습 모델의 구조를 혼합하여 항응고제 처방용량을 예측하는 흐름을 나타내는 모식도이다.

도9는 일 실시예에 따른 제1 기계학습 모델 및 제2 기계학습 모델에 적용되는 트레이닝 데이터를 전처리하는 예시도이다.

도10은 일 실시예에 따른 제1 기계학습 모델의 구조와 제2 기계학습 모델의 구조를 혼합하여 각각의 모델의 가중치를 조정하는 흐름을 나타내는 모식도이다.

발명을 실시하기 위한 구체적인 내용

- [0032] 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.
- [0033] 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.
- [0034] 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.
- [0035] 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.
- [0036] 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.
- [0037] 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.
- [0038] 경구용 항응고제 중 하나인 와파린의 용량을 결정하는 방법은 예시적으로 혈액응고검사(PT INR)를 사용할 수 있다. 이는 혈액응고검사를 수시로 모니터링하여 개인별 용량을 결정하는 방법으로 적정용량을 결정하기 위해 수 개월이 소요될 수 있다. 또한, 적정용량을 결정하기 위해 와파린이 과다 투여되거나 과소 투여되는 경우, 혈전 색전증 또는 출혈의 위험에 놓이게 되는 문제가 발생할 수 있다.
- [0039] 한편, 와파린 용량에 영향을 미치는 인자는 유전적(genetic) 요인(factor), 환경적(environmental) 요인, 클리닉(clinical) 요인 등으로 다양할 수 있다. 그 중에 유전적 요인은 예시적으로, 40%로 가장 높은 영향을 미치는 비중을 가지고 있을 수 있다. 유전적 요인은 각 환자가 가진 유전형(genotype)의 차이에 따라 와파린 대사 정도가 달라지는 특징을 가질 수 있다. 이에 따라 유전형 검사는 최적의 와파린 용량을 결정하는 또 다른 연구들에 포함될 수 있다. 그러나 유전형 검사는 대용량의 유전자데이터를 저장하는 데이터베이스가 구축되어야 하는 문제가 있을 수 있다. 또한, 유전형 검사를 활용하는 지침도 미흡한 상태인 문제 또한 있을 수 있다. 즉, 와파린 용량 결정을 위한 유전형 검사는 비용 및 시간이 효율적이지 않는 문제가 있을 수 있다. 이에 따라 유전형 검사 연구를 실제 병원 현장에서 적용하기에는 다소 어려움이 따를 수 있다.

- [0040] 이에 본 발명은 상기 문제점을 해결하기 위해 개발한 기계학습 모델을 통해 클리닉 데이터(clinical data)로만 최적의 와파린 용량을 예측할 수 있다.
- [0041] 우선, 도1을 들어 기계학습 모델을 이용하여 항응고제 처방용량을 예측하는 전자 장치에 대해 상세하게 설명하도록 한다.
- [0042] 도1은 일 실시예에 따른 항응고제 처방용량 예측 장치의 예시적인 구조를 도시한다.
- [0043] 일 실시예에 따른 장치(예: 전자 장치(100))는 환자 입원 시 수신되는 환자의 클리닉 데이터에 근거하여 예측되는 항응고제 처방용량을 사용자(예: 의사)에게 제공할 수 있다. 참고로 본 명세서에서 항응고제가 와파린인 예시를 주로 설명하나, 이로 한정하는 것은 아니다. 다른 항응고제에 대해서도 후술하는 제1 기계학습 모델 및 제2 기계학습 모델이 구축 및 트레이닝되어, 예측에 사용될 수도 있다.
- [0044] 전자 장치(100)는 프로세서(110) 및 메모리(120)를 포함할 수 있다.
- [0045] 프로세서(110)는 환자의 클리닉 데이터에 기초하여 입력 데이터를 생성할 수 있다. 프로세서(110)는 메모리(120)에 저장된 제1 기계학습 모델에 입력 데이터를 적용함으로써 복수의 처방용량 구간들에 대한 복수의 예측 결과들을 출력할 수 있다. 제1 기계학습 모델은 서브 기계학습 모델들을 포함할 수 있다. 프로세서(110)는 입력 데이터에 각 서브 기계학습 모델을 적용함으로써 해당하는 처방용량 구간에 대한 예측 결과를 출력할 수 있다.
- [0046] 서브 기계학습 모델들은 항응고제의 처방 예상 범위가 단위 용량으로 분할된 복수의 처방용량 구간들의 각각에 대응될 수 있다. 처방 예상 범위는 해당하는 항응고제에 대해 의료진(예: 의사 등)에 의해 주로 처방될 것으로 예상되는 범위를 나타낼 수 있다. 예를 들어, 처방 예상 범위는 해당 항응고제가 유효한 임상 효과를 나타내는 용량의 범위일 수 있고, 와파린의 경우 0.5mg 이상 10.0mg 미만일 수 있다. 단위 용량이 0.5mg인 경우, 복수의 처방용량 구간들은 1.0mg 이상 1.5mg 미만, 내지 9.5mg 이상 10.0mg미만으로 분할될 수 있다. 다만, 복수의 처방용량 구간들의 분할을 전술한 바로 한정하는 것은 아니다.
- [0047] 프로세서(110)는 서브 기계학습 모델들의 각각에서 해당하는 처방용량 구간에서의 처방 여부를 지시하는 출력을 생성할 수 있다. 예를 들어, 서브 기계학습 모델들 중 한 서브 기계학습 모델은 한 처방용량 구간(예: 1mg 이상 1.5mg 미만의 구간)에 대응할 수 있다. 해당 서브 기계학습 모델의 출력이 참(예: 1의 비트 값)인 경우, 해당 출력은 적어도 해당 처방용량 구간의 용량이 처방되어야 한다는 것을 나타낼 수 있다. 예를 들어, 1mg 이상 1.5mg 미만의 구간의 처방용량 구간에 대응하는 서브 기계학습 모델의 출력이 참인 경우, 해당 출력은 적어도 1mg의 항응고제가 처방되어야 한다는 추정 결과를 나타낼 수 있다.
- [0048] 프로세서(110)는 전술한 바와 같이 출력된 복수의 예측 결과들인 생성된 출력 중 해당하는 처방용량 구간에서 처방하는 것으로 결정된 출력의 개수에 대응하여 환자의 항응고제 처방용량을 예측할 수 있다. 아래 도 2에서는 서브 기계학습 모델들의 출력들을 종합적으로 고려하여 항응고제의 처방용량을 예측하는 동작이 설명된다.
- [0049] 참고로, 프로세서(110)는, 제2 기계학습 모델이 추가되는 경우, 제1 기계학습 모델에서 결정된 항응고제의 처방용량과 제2 기계학습 모델에서 결정된 항응고제의 처방용량에 기초하여 최종 항응고제의 처방용량을 결정할 수 있다. 이러한 항응고제 처방용량을 예측하는 구체적인 방식은 도 4 내지 도8에서 후술한다.
- [0050] 트레이닝 과정에서, 프로세서(110)는 트레이닝 출력의 참값용량과 제1 기계학습 모델 및 제2 기계학습 모델에서 산출한 항응고제 처방용량을 각각 비교하여 로스를 결정할 수 있다. 또한, 각각의 로스가 임계 범위를 초과하는 경우, 제1 기계학습 모델 및 제2 기계학습 모델 내에서의 각각의 가중치를 조정된 후 항응고제 처방용량을 각각 재산출하여 로스를 재결정할 수 있다. 로스 결정과 같은 항응고제 처방용량 검증작업은 도10에서 구체적으로 후술한다. 본 명세서에서는 전자 장치(100)가 트레이닝 동작 및 처방용량의 예측 동작을 둘 다 수행하는 것으로 설명하였으나, 이로 한정하는 것은 아니다. 트레이닝 동작은 별도 장치에 의해 미리 수행될 수도 있다.
- [0051] 메모리(120)는 제1 기계학습 모델을 저장할 수 있다. 전술한 바와 같이 제1 기계학습 모델은 복수의 서브 기계학습 모델을 포함할 수 있다. 제1 기계학습 모델은 복수의 서브 기계학습 모델들의 출력들에 기초하여 예측된 처방용량을 출력하도록 설계 및 트레이닝된 모델로서, 예시적으로, 뉴럴 네트워크(예: 컨볼루션 뉴럴 네트워크(CNN, convolutional neural network))를 포함할 수 있다. 후술하겠으나, 제1 기계학습 모델은 랭킹 CNN에 기초한 모델일 수 있다. 서브 기계학습 모델은 처방 예상 범위의 처방용량 구간 별로 준비될 수 있다. 각 서브 기계학습 모델은 입력 데이터로부터 해당하는 처방용량 구간의 용량이 적어도 처방되어야 하는지 여부를 지시하는 값을 출력하도록 설계 및 트레이닝될 수 있다.

- [0052] 추가로, 메모리(120)는 제2 기계학습 모델도 저장할 수 있다. 제2 기계학습 모델은 입력 데이터로부터 추정된 처방용량을 지시하는 값(예: 연속된 실수 값)을 출력하도록 설계 및 트레이닝될 수 있다. 제2 기계학습 모델의 출력 값은 전술한 처방 예상 범위 내의 값으로 제한될 수 있으나 이로 한정하는 것은 아니다.
- [0053] 메모리(120)는 트레이닝 데이터를 저장할 수 있다. 트레이닝 데이터는 클리닉 데이터 중으로부터 추출된 일부 데이터를 포함하는 트레이닝 입력 및 환자의 퇴원 시 의사에 의해 처방된 항응고제 처방용량인 참값 용량을 라벨로 가지는 트레이닝 출력의 쌍을 포함할 수 있다. 트레이닝 데이터는 전술한 제1 기계학습 모델 및 제2 기계학습 모델의 트레이닝을 위해 사용될 수 있다. 메모리(120)는 클리닉 데이터, 항응고제 처방을 위한 입력 데이터 및 예측된 항응고제 처방용량 또한 저장할 수 있다. 여기서 입력 데이터는 클리닉 데이터에 기초하여 인구통계학적정보, 활력징후, 진단정보 및 복약정보를 포함하는데, 입력 데이터의 더욱 상세한 사항은 하기 도 2에서 후술한다. 메모리(120)는 제1 기계학습 모델 및 제2 기계학습 모델과 항응고제 처방을 위한 입력 데이터, 예측된 항응고제 처방용량 및 학습을 위한 트레이닝 데이터를 각각 별도의 저장공간으로 분리하여 저장할 수 있다. 또한, 메모리(120)는 항응고제 처방을 위하여 클리닉 데이터를 입력 데이터로 생성하는 프로그램 및 학습을 위하여 클리닉 데이터를 트레이닝 데이터로 추출하는 프로그램 또한 저장할 수 있다.
- [0054] 또한, 전자 장치(100)는 통신부를 더 포함할 수 있다. 통신부는 유선 또는 무선 네트워크를 통해 다른 장치와 통신을 수립하고, 정보를 수신하고 전송하는 통신 회로를 포함할 수 있다. 통신부는 외부 장치로부터 환자의 클리닉 데이터를 수신할 수 있다. 프로세서(110)는, 통신부를 통해 수신된 환자의 클리닉 데이터를 기계학습 모델에 적용될 수 있는 입력 데이터로 생성할 수 있다. 다른 예를 들어, 통신부는 기계학습 모델의 학습을 위한 클리닉 데이터도 수신할 수 있다. 프로세서(110)는 통신부를 통해 수신된 학습을 위한 클리닉 데이터로부터 트레이닝 데이터를 추출할 수 있다.
- [0055] 통신부는 항응고제 처방용량을 외부 장치에게 송신할 수 있다. 다만, 이로 한정하는 것은 아니고, 전자 장치(100)는 예측 및/또는 추정된 처방용량을 전자 장치(100) 내부에 포함된 디스플레이(미도시됨)를 통해 출력할 수도 있다.
- [0056] 도2는 일 실시예에 따른 항응고제 처방용량 예측 방법의 흐름도이다.
- [0057] 일 실시 예에 따르면, 아래의 단계들(210 내지 240)은 장치(예: 도1의 전자 장치(100))에 의해 수행될 수 있다.
- [0058] 단계(210)에서, 전자 장치의 통신부는 환자의 클리닉 데이터를 수신할 수 있다. 여기서의 클리닉 데이터는 항응고제 처방용량을 예측하기 위한 클리닉 데이터에 해당될 수 있다.
- [0059] 단계(220)에서, 전자 장치의 프로세서는 클리닉 데이터에 기초하여 입력 데이터를 생성할 수 있다. 프로세서는 클리닉 데이터로부터 인구통계학적정보(demographic information), 활력징후(Vital sign)정보, 진단정보 및 복약정보를 포함하는 입력 데이터를 생성할 수 있다. 프로세서는 인구통계학적 정보, 활력징후 정보, 진단 정보 및 복약 정보 별로 복수의 특징 변수들을 추출할 수 있다. 예를 들어, 인구통계학적 정보(demographic information)로부터 4개의 특징 변수, 활력징후(Vital sign) 정보로부터 6개의 특징 변수, 진단정보로부터 18개의 특징 변수, 투약정보로부터 20개의 특징 변수들을 추출할 수 있다. 프로세서는 클리닉 데이터 중 총 48개의 특징 변수를 변수로 가지는 입력 데이터를 생성할 수 있다. 아래에서는 특징 변수의 예시적인 항목을 설명한다.
- [0060] 단계(230)에서, 프로세서는 복수의 예측 결과들을 출력할 수 있다. 예를 들어, 프로세서는 서브 기계학습 모델들의 각각에서 해당하는 처방용량 구간에서의 처방 여부를 지시하는 출력을 생성할 수 있다. 전술한 바와 같이, 프로세서는 입력 데이터에 복수의 서브 기계학습 모델들을 적용함으로써 복수의 예측 결과들을 생성할 수 있다.
- [0061] 항응고제의 처방 예상 범위는, 단위 용량 구간에 기초하여 분할된, 복수의 처방용량 구간들로 형성될 수 있다. 복수의 처방용량 구간들은 단위 용량 구간에 기초하여 등간격으로 구분될 수 있다. 예를 들어, 복수의 처방용량 구간들의 각각의 하한치들 간의 간격은 모두 단위 용량 구간(예: 0.5mg)만큼일 수 있다. 서브 기계학습 모델들의 각각은 복수의 처방용량 구간들의 각각에 개별적으로 대응할 수 있다. 서브 기계학습 모델들은 제1 기계학습 모델에 포함될 수 있다. 제1 기계학습 모델은 입력 데이터가 적용되어 복수의 처방용량 구간들에 대한 복수의 예측 결과들을 출력할 수 있다. 복수의 예측 결과들은 각각 이진수의 형태로 출력될 수 있다. 이러한 제1 기계학습 모델의 구조는 도4를 들어 좀 더 상세하게 설명한다.
- [0062] 단계(240)에서, 전자 장치의 프로세서는 출력된 복수의 예측 결과들에 기초하여 환자의 항응고제 처방용량을 예

측할 수 있다. 처방용량은 서브 기계학습 모델들로부터 생성된 출력 중 해당하는 처방용량 구간에서 처방하는 것으로 결정된 출력의 개수에 대응하여 예측될 수 있다. 서브 기계학습 모델들로부터 생성되어 출력된 복수의 예측 결과들로부터 처방용량을 예측하는 방법은 도5 및 도6을 들어 더욱 상세하게 설명한다.

[0063] 도3은 일 실시예에 따른 클리닉 데이터 전체항목을 SHAP 기법을 이용하여 특징 변수를 시각화한 예시도이다.

[0064] 특징 변수는 SHAP(SHapley Additive exPlanations) 기법에 따라 선정될 수 있다. SHAP 기법은 특징 변수의 중요도를 검증하는 기법이다. SHAP 기법은 중요도를 시각화하여 확인할 수 있도록 함에 따라, 클리닉 데이터에서 의미있다고 판단되는 특징 변수를 용이하게 선별하는데 사용자에게 도움을 줄 수 있다. 항응고제의 처방용량을 예측하기 위하여 클리닉 데이터의 전체 항목을 모두 고려하는 것은 실제 병원 현장에서 적용되기 어렵기 때문에 이 작업은 상당히 유용할 수 있다. 도3에서 도시되는 높은 중요도를 나타내는 특징 변수는 환자의 진단시점 나이, 체중, 특정 질환들의 진단 여부, 산소포화도, 신장측정 값, 수축기 혈압 등일 수 있다. 여기서 와파린 용량의 결정과 임상적으로 의미있다고 판단되는 특징 변수는 예를 들어 48개로 선정될 수 있다.

[0065] 클리닉 데이터 중 입력 데이터에 포함되는 특징변수들의 예시는 다음과 같다.

[0066] 인구통계학적 정보는 성별, 나이, 신장, 체중을 포함할 수 있다.

[0067] 활력징후는 맥박수, 산소포화도, 수축기 혈압, 이완기 혈압, 체온, 호흡수를 포함할 수 있다.

[0068] 진단정보는 협심증(Angina), 부정맥(Arrhythmia), 심박세동(Atrial fibrillation), 암(Cancer), 만성 허혈성 심장병(Chronic ischemic heart disease), 만성폐질환(Chronic lung disease), 진성 당뇨병(Diabetes mellitus), 이상지질혈증(Dyslipidemia), 심부전(Heart failure), 고혈압(Hypertension), 두개 강 내 출혈(Intracranial bleeding), 간 질환(Liver disease), 심근경색(Myocardial infarction), 말초동맥질환(Peripheral arterial disease), 폐색전증(Pulmonary embolism), 신장질환(Renal disease), 뇌졸중/일과성 허혈발작(Stroke/TIA), 판막성심장병(Valvular heart disease)을 포함할 수 있다.

[0069] 투약정보는 ACE 억제제(ACE inhibitor), ADP 수용체 억제제(ADP receptor inhibitor), 안지오텐신 II 수용체차단제(ARB), 알도스테론 길항제(Aldosterone antagonist), 알로푸리놀(Allopurinol), 아스피린(Aspirin), 베타 차단제(Beta-blocker), 칼슘통로차단제(Calcium channel blocker)- 디하이드로피리딘(dihydropyridine), 칼슘통로차단제(Calcium channel blocker)- 비디하이드로피리딘(non-dihydropyridine), DPP4 억제제(DPP4 inhibitor), 항응고제(Direct oral anticoagulant), 이뇨제, 고리(Diuretic, loop), 이뇨제, 타이아자이드(Diuretic, thiazide), 인슐린(Insulin), 메트포르민(Metformin), 질산염(Nitrate), 다른 항협심증제(Other antianginal), 다른 지질 강하(Other lipid lowering), 혈관 내 콜레스테롤 억제제(Statin), 설폰닐유레아(Sulfonylurea)를 포함할 수 있다.

[0070] 도4는 일 실시예에 따른 제1 기계학습 모델의 구조에서 복수의 예측 결과들이 출력되는 흐름을 나타내는 모식도이다. 도5는 일 실시예에 따른 제1 기계학습 모델의 구조에서 항응고제 처방용량을 예측하는 예시도이다. 도6은 일 실시예에 따른 제1 기계학습 모델의 구조에서 항응고제 처방용량을 예측하는 또 다른 예시도이다.

[0071] 제1 기계학습 모델 구조는 랭킹 CNN(Ranking-Convolutional Neural Networks)에 기초한 구조일 수 있다. 처방 예상 범위가, 예를 들어, 10mg으로 총 19개의 처방용량 구간들로 형성될 수 있다. 19개의 처방용량 구간들의 각각 별로 서브 기계학습 모델이 준비될 수 있다. X에 해당하는 입력 데이터(410)가 들어오면 각 구간에 해당하는 서브 기계학습 모델들(420)의 각각(420_1, 420_2, ... 420_19)은 처방 여부를 지시하는 출력(430)을 생성할 수 있다. 서브 기계학습 모델들이 19개이므로, 19개의 처방용량 구간들에 대한 예측 결과들이 출력될 수 있다.

[0072] 19개의 각 구간들의 서브 기계학습 모델은 해당 구간에서 용량이 처방되어야 하는 확률을 출력할 수 있다(431). 예를 들어, 구간1에서의 서브 기계학습 모델에서 산출된 결과가 용량이 처방되어야 할 확률에 포함되는 경우, 프로세서는 1의 값으로 출력할 수 있다. 만약 서브 기계학습 모델에서 산출된 결과가 용량이 처방되어야 할 확률에 포함되지 않는 경우, 프로세서는 0의 값으로 출력할 수 있다. 프로세서는 산출된 확률 값이 임계 확률(예: 1)을 초과하는 경우 해당 구간에 대해 1의 값을 출력할 수 있다. 프로세서는 산출된 확률 값이 임계 확률 이하인 경우 해당 구간에 대해 0의 값을 출력할 수 있다. 다만, 임계 확률은 예시적인 것으로, 이로 한정하는 것은 아니다. 프로세서는 각각 출력된 1 값의 합에 대응하는 값(예: 범위)으로 와파린 용량을 예측할 수 있다. 1의 값이 출력된 개수가 k인 경우, 예측된 와파린 용량은 ((k+1)×단위 용량 구간) 미만일 수 있다. (즉, 예측된 와파린 용량 < (k+1)×단위 용량 구간) k는 0이상의 정수일 수 있다. 처방 예상 범위가 P, 처방용량 구간들의 개수가 M인 경우, 단위 용량 구간은 P/(M+1)일 수 있다. P는 실수(예를 들어, mg 단위), M은 2이

상의 정수일 수 있다. 예를 들어, 도5에서는 1 내지 17 구간인 총 17개 구간(예: $k=17$)이 1(510)의 값으로 출력되었고 18 및 19 구간에서는 0(520)의 값으로 출력된 것으로 볼 수 있다(432). 이에 따라 최적 와파린 용량은 17개 개수에 따라 9mg (즉, $(17+1) \times 0.5\text{mg}$) 미만일 것으로 예측할 수 있다. 도6에서는 구간 1 내지 5, 7 내지 16, 18 및 19에서 1(610)의 값으로 출력되었고, 구간 6 및 17에서는 0(620)의 값으로 출력된 것으로 볼 수 있다. 이에 따라 최적 와파린 용량은 17개 개수에 따라 9mg 미만일 것으로 예측할 수 있다. 즉, 도5와 도6에서 각각의 구간 결과 값은 다르게 출력되었지만 최적 와파린 용량은 동일하게 예측되는 것을 볼 수 있다.

[0073] 항응고제 처방용량은 제1 기계학습 모델 구조에 제2 기계학습 모델 구조를 추가하여 예측할 수 있다. 제2 기계학습 모델 구조가 추가된 형태는 도7 및 도8을 들어 더욱 상세하게 설명한다.

[0074] 도7은 일 실시예에 따른 제2 기계학습 모델의 구조에서 항응고제 처방용량을 예측하는 흐름을 나타내는 모식도이다. 도8은 일 실시예에 따른 제1 기계학습 모델의 구조와 제2 기계학습 모델의 구조를 혼합하여 항응고제 처방용량을 예측하는 흐름을 나타내는 모식도이다.

[0075] 제2 기계학습 모델 구조는 선형회귀분석(Linear Regression)에 기초한 기계학습 모델로서, 예를 들어, 뉴럴 네트워크를 포함할 수 있다. 제2 기계학습 모델 구조는 제1 기계학습 모델 구조에서 단위 용량 구간의 분할없이 처방용량을 예측할 수 있다. 즉, X에 해당하는 입력 데이터(710)가 들어오면 제2 기계학습 모델(720)은 처방 여부를 지시하는 출력(730)을 생성할 수 있다. 도8을 참조하면, 프로세서는 X에 해당하는 입력 데이터(810)가 수신되면 제1 기계학습 모델 구조(821)에서 산출한 항응고제 처방용량(831)과 제2 기계학습 모델 구조(822)에서 산출한 항응고제 처방용량(832)에 기초하여 최종 처방용량(850)을 결정할 수 있다. 최종 처방용량은 제1 기계학습 모델 구조의 처방용량과 제2 기계학습 모델 구조의 처방용량의 평균으로 산출(840)될 수 있다. 제2 기계학습 모델 구조가 추가된 구조에 XGBoost에 기초한 제3 기계학습 모델 구조도 더 추가될 수 있다.

[0076] 제1 기계학습 모델 및 제2 기계학습 모델을 트레이닝하기 위하여 데이터를 전처리하는 방식에 대해 도9를 들어 설명한다.

[0077] 도9는 일 실시예에 따른 제1 기계학습 모델 및 제2 기계학습 모델에 적용되는 트레이닝 데이터를 전처리하는 예시도이다.

[0078] 제1 기계학습 모델 및 제2 기계학습 모델은 클리닉 데이터 중으로부터 추출된 일부 데이터를 포함하는 트레이닝 입력 및 참값(ground truth) 용량을 라벨로 가지는 트레이닝 출력의 쌍을 포함하는 트레이닝 데이터를 이용하여 트레이닝될 수 있다. 클리닉 데이터를 추출하는 기준은 도9에서 예시한 바와 같이 포함기준에서 미포함되는 기준을 제외시키는 방식을 반복하여 수행한다. 예를 들어, 와파린을 처방한 19세 이상 CV 또는 CS 입원 건을 4,079건이라고 할 수 있다. 이에 제1 미포함기준(910)을 와파린 퇴원 처방이 없는 건으로 적용하여 4,079 건에서 337건을 제외시킬 수 있다. 즉, 와파린을 처방한 19세 이상 CV 또는 CS 입원 건 중 와파린 퇴원 처방이 있는 입원 건은 3,742 건일 수 있다. 여기서 제2 미포함기준(920)을 와파린 처방이 3회 미만인 입원 건으로 적용하여 3,742건에서 344건을 제외시킬 수 있다. 즉, 와파린을 처방한 19세 이상 CV 또는 CS 입원 건 중 와파린 퇴원 처방이 있으면서 와파린 처방이 3회 이상인 건은 3,398건일 수 있다. 여기에 제3 미포함기준(930)을 입원 1,2일에 체중값이 없는 입원 건으로 적용하여 3,398건에서 230건을 제외시킬 수 있다. 즉, 와파린 처방이 3회 이상인 19세 이상 CV 또는 CS 입원 건 중 와파린 퇴원 처방이 있으면서 입원 1,2일에 체중값이 있는 입원 건은 3,168건일 수 있다. 이와 같은 데이터에 미포함 기준을 추가하여 트레이닝 데이터를 선별할 수 있다. 제1 기계학습 모델 및 제2 기계학습 모델의 트레이닝은 항응고제 처방용량을 예측하기 사전에 수행되어야 한다.

[0079] 트레이닝 데이터에 포함된 참값(ground truth)용량이란 환자 퇴원 시 의사에 의해 처음으로 처방된 항응고제 처방용량이 라벨링된 것을 의미할 수 있다. 참값용량은 제1 기계학습 모델에서 산출한 항응고제 처방용량과 비교될 수 있으며, 이러한 비교를 통하여 모델의 가중치를 조정하는 방식은 도10을 들어 상세하게 설명한다.

[0080] 도10은 일 실시예에 따른 제1 기계학습 모델의 구조와 제2 기계학습 모델의 구조를 혼합하여 각각의 모델의 가중치를 조정하는 흐름을 나타내는 모식도이다.

[0081] 프로세서는 수신된 X에 해당하는 입력 데이터(1010)를 제1 기계학습 모델 구조(1021)에 적용할 수 있다. 프로세서는 제1 기계학습 모델에서 산출한 처방용량(1031)과 참값용량(1032)을 비교하여 발생한 차이 값을 로스(1041)로 결정할 수 있다. 프로세서는 로스(1041)를 임계 범위와 비교할 수 있다. 로스(1041)가 임계 범위 이하인 경우, 제1 기계학습 모델(1021)의 가중치에는 이상이 없는 것으로 검증할 수 있다. 그러나 로스(1041)가 임계 범위를 초과하는 경우, 제1 기계학습 모델(1021)의 가중치를 조정할 수 있다. 프로세서는 로스(1041)가 임계 범위 내로 될 때까지 로스의 결정과, 가중치의 조정을 반복할 수 있다.

[0082] 프로세서는 제1 기계학습 모델과 다른 구조를 가지는 제2 기계학습 모델 또한 트레이닝 데이터를 이용하여 트레이닝할 수 있다. 제1 기계학습 모델의 검증방식과 동일하게 프로세서는 제2 기계학습 모델(1022)에서 산출한 항응고제 처방용량(1033)을 참값용량(1032)과 비교하여 로스(1042)를 산출할 수 있다. 프로세서는 산출된 로스(1042)가 임계 범위 내로 될 때까지 로스의 결과와, 가중치의 조정을 반복할 수 있다.

[0083] 참고로, 전체 클리닉 데이터 중 80%의 데이터가 트레이닝 데이터로 사용될 수 있고, 전체 클리닉 데이터 중 20%의 데이터는 전술한 바와 같이 트레이닝된 모델을 평가하는 테스트 데이터로 사용될 수 있다. 트레이닝 데이터로 각 훈련된 모델들은 예시적으로 표1에 표기된 바와 같이 절대평균오차(MAE, Mean Absolute Error)로 평가할 수 있다. 절대평균오차는 실제 값과 예측 값 간의 오차(Error)를 절대값으로 변환하여 평균화시켰으므로 값이 클수록 오차가 큰 것으로 볼 수 있다. 즉, 의사가 최초 입원 시 처방한 용량이 제일 오차가 큰 것으로 볼 수 있으며, 이에 비해 기계학습 모델의 오차들은 상대적으로 비슷한 수준의 낮은 값을 나타내는 것으로 볼 수 있다.

표 1

모델	MAE(mg)
의사(최초 처방용량)	1.213
제1 기계학습 모델	0.935
제2 기계학습 모델	0.925
제3 기계학습 모델	0.880

[0085] 이와 같은 결과는 정확도를 표시한 표2를 통해서도 확인할 수 있다. 표2에서 기계학습 모델을 통해 산출된 정확도는 전부 의사가 최초 처방한 용량에 비해 높은 정확도를 나타내는 것으로 볼 수 있다.

표 2

모델	정확도 (e=1.0mg) (%)	정확도 (e=1.5mg) (%)	정확도 (e=2.0mg) (%)
의사(최초 처방용량)	58.83	69.09	85.49
제1 기계학습 모델	61.83	80.44	91.01
제2 기계학습 모델	63.72	79.65	92.11
제3 기계학습 모델	65.46	80.60	89.43

[0087] 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, 콘트롤러, 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 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.

[0088] 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(embodiment)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.

[0089] 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.

[0090] 위에서 설명한 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 또는 복수의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

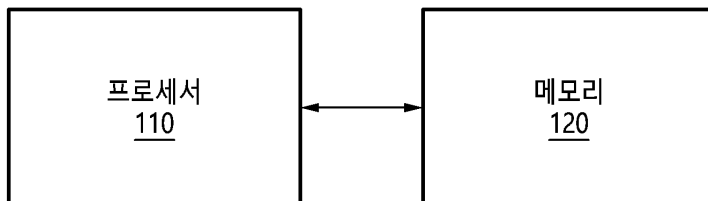
[0091] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 이를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[0092] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

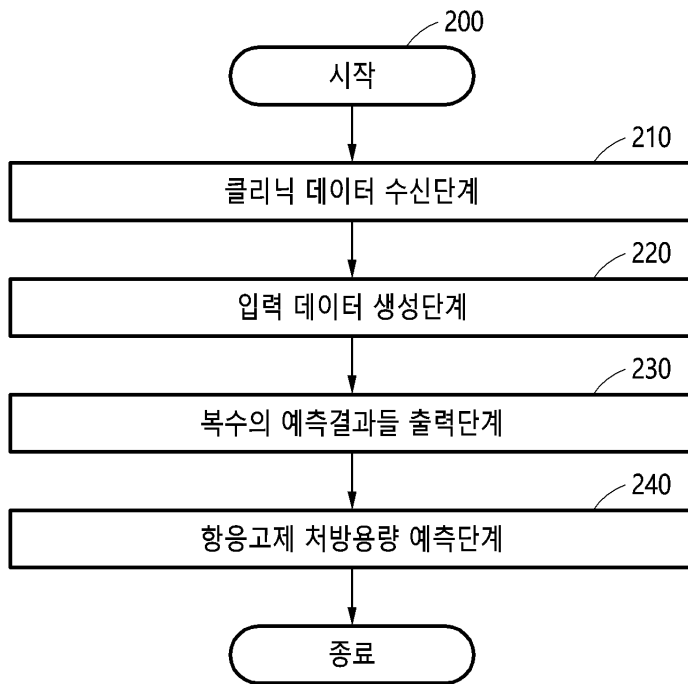
도면

도면1

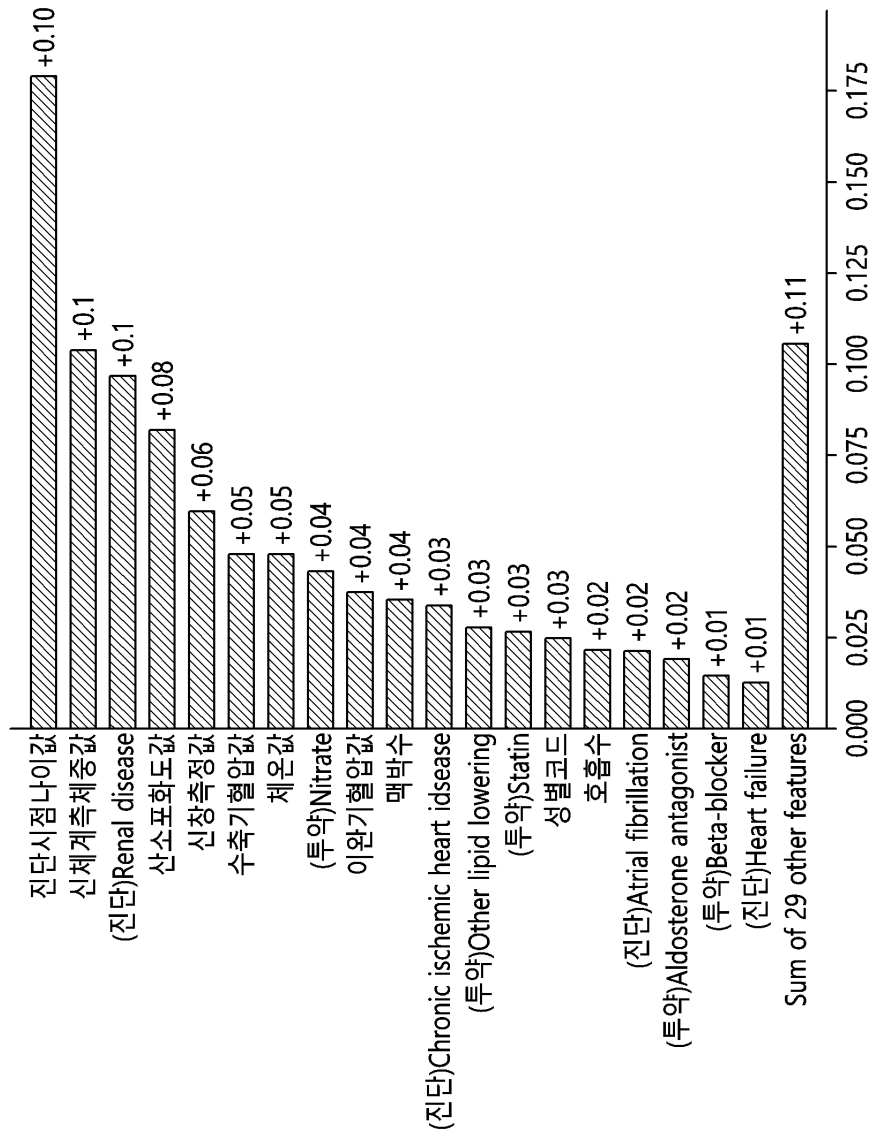
100



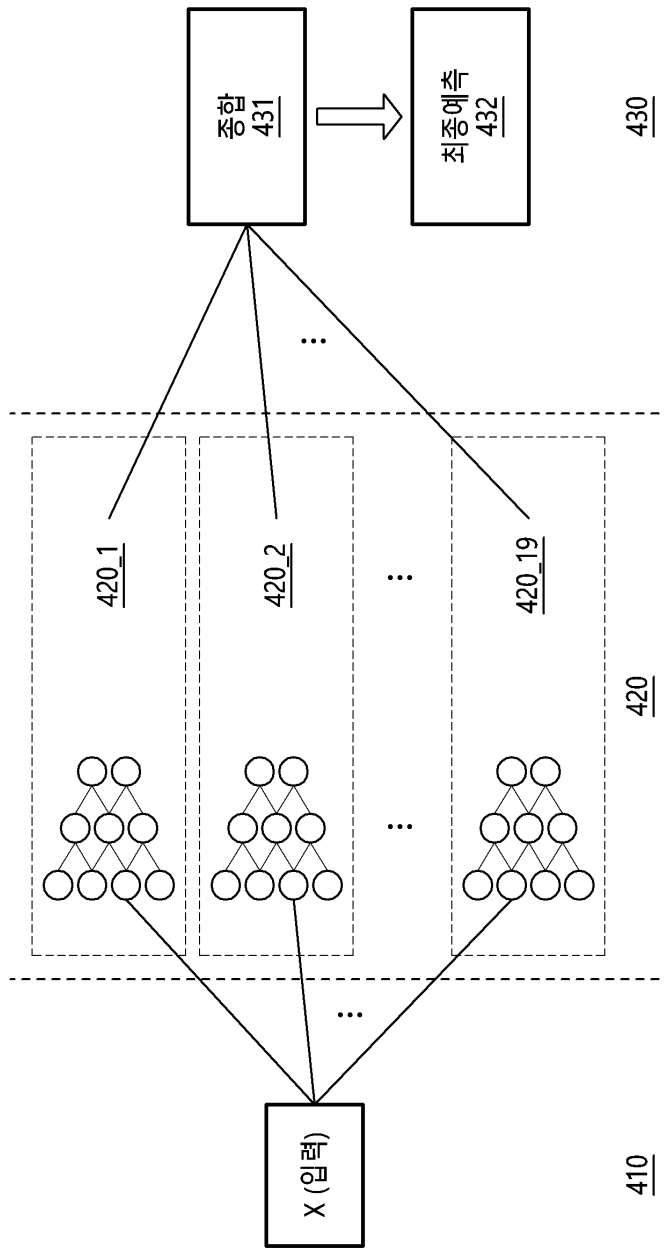
도면2



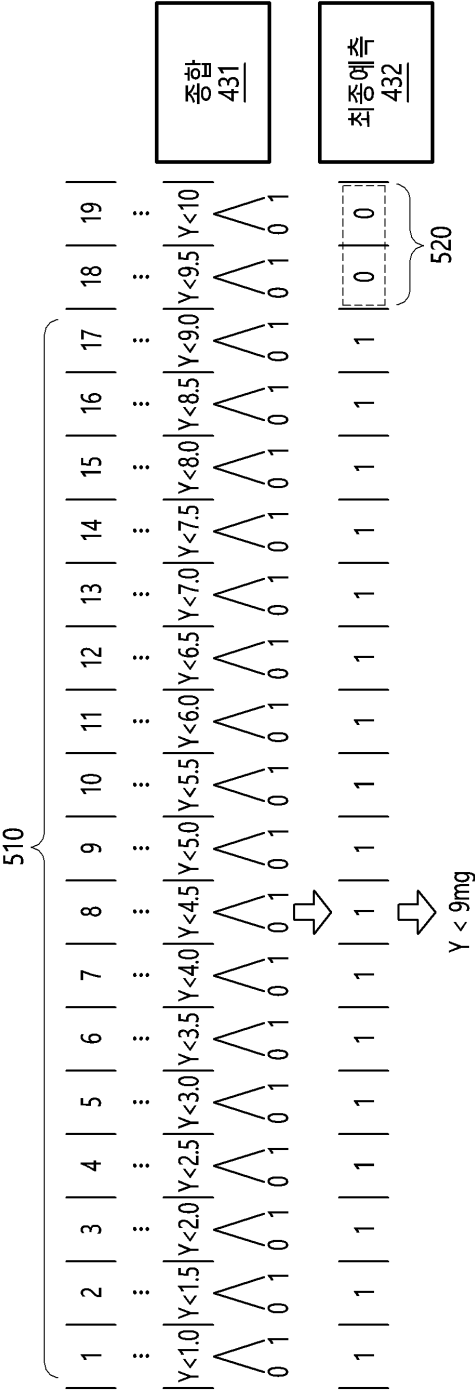
도면3



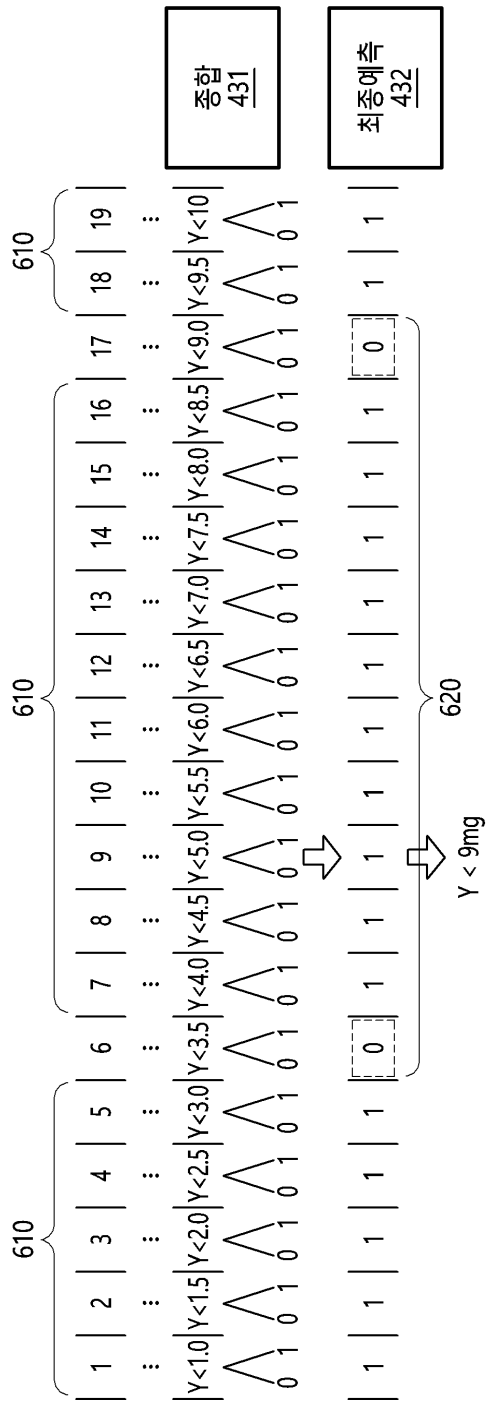
도면4



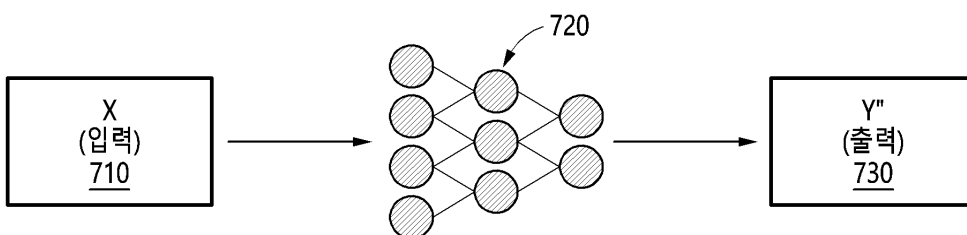
도면5



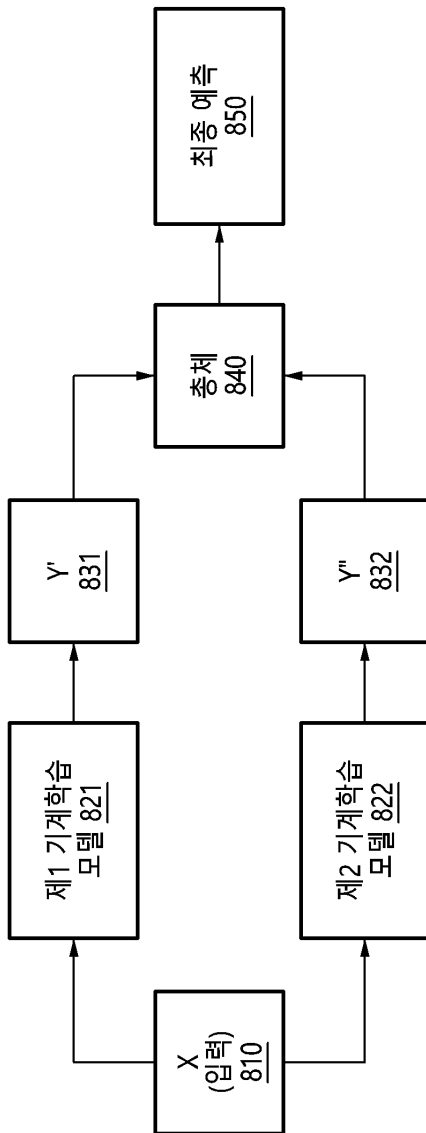
도면6



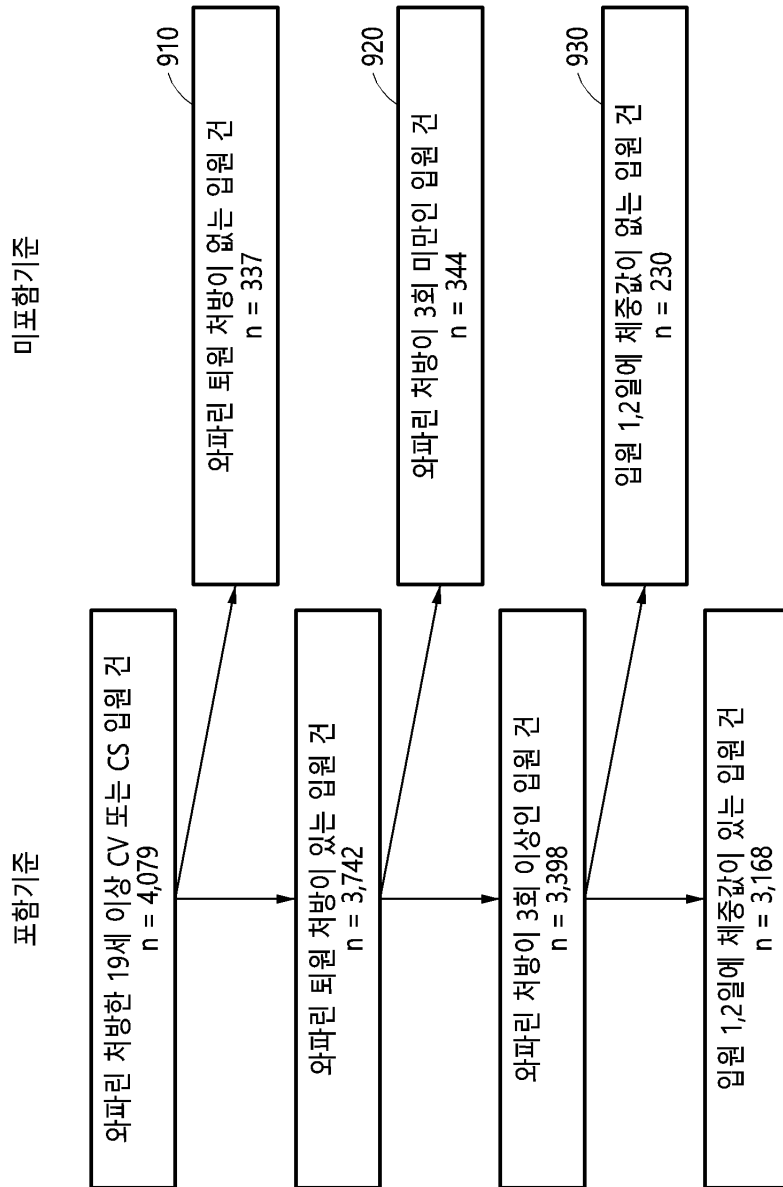
도면7



도면8



도면9



도면10

