(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0035663
(43) 공개일자 2024년03월18일

(51) 국제특허분류(Int. Cl.)
G06N 20/20 (2019.01) *G06N 3/04* (2023.01)
G06N 3/08 (2023.01) *G16H 10/60* (2018.01)
G16H 50/70 (2018.01)
(52) CPC특허분류
G06N 20/20 (2021.08)
G06N 3/042 (2023.01)
(21) 출원번호 10-2022-0114113
(22) 출원일자 2022년09월08일
심사청구일자 2022년09월08일

(71) 출원인
재단법인 아산사회복지재단
서울특별시 송파구 올림픽로43길 88 (풍납동)
울산대학교 산학협력단
울산광역시 남구 대학로 93(무거동)
(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

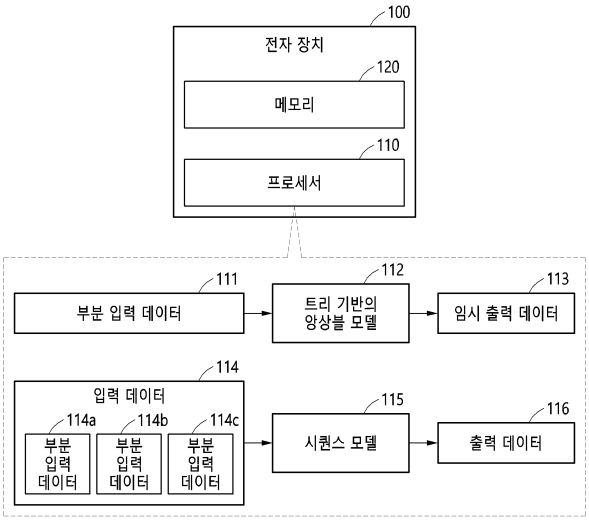
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **트리 기반의 앙상블 모델을 이용하여 시퀀스 모델의 입력 데이터의 피처를 선택하는 방법 및 장치**

(57) 요 약

전자 장치에 있어서, 의료 데이터의 복수의 피처들(features) 중에서 트리 기반의 앙상블 모델(tree-based ensemble model)에 기초하여 선택된 피처를 가지는 부분 입력 데이터(partial input data)를 생성하고, 복수의 부분 입력 데이터들을 포함하는 입력 데이터(input data)를 시퀀스 데이터(sequence data)로 생성하며, 상기 생성된 시퀀스 데이터에 시퀀스 모델(sequence model)을 적용함으로써 출력 데이터(output data)를 산출하는 프로세서를 포함할 수 있다.

대 표 도 - 도1



(52) CPC특허분류
G06N 3/045 (2023.01)
G06N 3/08 (2023.01)
G16H 10/60 (2021.08)
G16H 50/70 (2018.01)
(72) 발명자
안임진
경기도 의정부시 진등로19번길 20, 901동 1207호(녹양동, 녹양동 신도9차 전원마을아파트)
강희준
경기도 성남시 중원구 둔촌대로 300, 1004동 1004호(하대원동, 아튼빌아파트)
권한슬
서울특별시 노원구 초안산로5길 22, 105동 202호(월계동, 대우아파트)
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)
최희정
서울특별시 강남구 봉은사로49길 21-10, 202호(논현동, 블루하우스(Blue house))
김민경
경기도 의정부시 오목로 36, 103동 1502호(용현동, 현대아파트)
한지예
서울특별시 양천구 목동동로 10, 1110동 601호(신정동, 목동신시가지아파트11단지)
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 1465036574
과제번호 HR20C0026020022
부처명 보건복지부
과제관리(전문)기관명 한국보건산업진흥원
연구사업명 연구중심병원 육성 R&D
연구과제명 MeDIA: 의료 빅데이터, 사물인터넷, 인공지능 플랫폼 구축을 통한 혁신형 의료기술 개발
기 여 율 1/2
과제수행기관명 서울아산병원
연구기간 2022.01.01 ~ 2022.12.31

명 세 서

청구범위

청구항 1

전자 장치에 의하여 수행되는 방법에 있어서,

의료 데이터의 복수의 피처들(features) 중에서 트리 기반의 앙상블 모델(tree-based ensemble model)에 기초하여 부분 입력 데이터(partial input data)의 피처를 선택하는 단계;

상기 선택된 피처를 가지는 복수의 부분 입력 데이터들을 포함하는 입력 데이터(input data)를 시퀀스 데이터(sequence data)로 생성하는 단계; 및

상기 생성된 입력 데이터에 시퀀스 모델(sequence model)을 적용함으로써 출력 데이터(output data)를 산출하는 단계

를 포함하는 방법.

청구항 2

제1항에 있어서,

상기 부분 입력 데이터의 피처를 선택하는 단계는,

상기 의료 데이터에 상기 트리 기반의 앙상블 모델을 적용하여 산출된 임시 출력 데이터와 참 값(ground truth)을 비교함으로써 상기 트리 기반의 앙상블 모델을 트레이닝시키는 단계; 및

상기 트레이닝된 트리 기반의 앙상블 모델에 대한 상기 복수의 피처들 각각의 피처 중요도에 기초하여, 상기 복수의 피처들 중 적어도 하나의 피처를 선택하는 단계를 포함하는,

방법.

청구항 3

제1항에 있어서,

상기 부분 입력 데이터의 피처를 선택하는 단계는,

트레이닝된 트리 기반의 앙상블 모델에 대한 피처 중요도에 기초하여, 상기 의료 데이터의 상기 복수의 피처들 중에서 미리 결정된 개수의 피처를 선택하는 단계를 포함하는,

방법.

청구항 4

제1항에 있어서,

상기 부분 입력 데이터의 피처를 선택하는 단계는,

상기 의료 데이터의 복수의 피처들 중 선택된 피처를 가지는 임시 부분 입력 데이터에 상기 트리 기반의 앙상블 모델을 적용하여 산출된 임시 출력 데이터와 참 값을 비교함으로써 상기 트리 기반의 앙상블 모델을 트레이닝시키는 단계;

상기 트레이닝된 트리 기반의 앙상블 모델에 대하여 계산된 상기 임시 부분 입력 데이터의 각 피처의 피처 중요도(feature importance)에 기초하여, 상기 임시 부분 입력 데이터의 피처 중 적어도 하나의 피처를 선택하는 단

계;

상기 선택된 적어도 하나의 피처를 가지는 데이터로 상기 임시 부분 입력 데이터를 업데이트하는 단계; 및

상기 업데이트된 임시 부분 입력 데이터에 적용됨으로써 임시 출력 데이터를 산출하는 앙상블 모델로 상기 트리 기반의 앙상블 모델을 업데이트하는 단계를 포함하는,

방법.

청구항 5

제1항에 있어서,

상기 부분 입력 데이터의 피처를 선택하는 단계는,

부분 입력 데이터의 피처들 간의 시간 순서와 독립적으로 임시 출력 데이터를 산출하는 상기 트리 기반의 앙상블 모델에 기초하여 피처를 선택하는 단계를 포함하는,

방법.

청구항 6

제1항에 있어서,

상기 입력 데이터를 생성하는 단계는,

상기 복수의 부분 입력 데이터들이 대응하는 시간대들의 시간 순서에 따라 복수의 부분 입력 데이터들을 결합(combine)함으로써 상기 입력 데이터를 생성하는 단계를 포함하는,

방법.

청구항 7

제1항에 있어서,

상기 출력 데이터를 산출하는 단계는,

상기 생성된 입력 데이터에 포함된 상기 복수의 부분 입력 데이터들의 결합 순서에 따라 다른 출력 데이터를 산출 가능한 상기 시퀀스 모델을 이용하여 상기 출력 데이터를 산출하는 단계를 포함하는,

방법.

청구항 8

제1항에 있어서,

상기 출력 데이터를 산출하는 단계는,

복수의 부분 입력 데이터들이 대응하는 시간대들 중 마지막 시간대의 종료 시각 이후의 출력 시간대에 대응하는 출력 데이터를 산출하는 단계를 포함하는

방법.

청구항 9

제1항에 있어서,

상기 부분 입력 데이터의 피처를 선택하는 단계는,

상기 트리 기반의 앙상블 모델로서 XGBoost, AdaBoost, GBM(Gradient Boosting Machine), LightGBM, CatBoost, 또는 histogram-based gradient boost 중 적어도 하나를 포함하는 부스팅 알고리즘(boosting algorithm) 기반의 모델에 기초하여 상기 부분 입력 데이터의 피처를 선택하는 단계를 포함하는,

방법.

청구항 10

제1항에 있어서,

상기 출력 데이터를 산출하는 단계는,

상기 시퀀스 모델로서, RNN(Recurrent Neural Network), LSTM(Long Short Term Memory), ARIMA(AutoregRessive Integrated Moving Average), Seq2Seq(Sequence to Sequence model), GRU(Gated Recurrent Unit), 또는 Transformer 중 적어도 하나를 포함하는 기계 학습 모델을 상기 입력 데이터에 적용함으로써 상기 출력 데이터를 산출하는 단계를 포함하는,

방법.

청구항 11

제1항에 있어서,

상기 출력 데이터를 산출하는 단계는,

추정 사구체 여과율(estimated Glomerular Filtration Rate; eGFR)을 포함하는 상기 출력 데이터를 산출하는 단계를 포함하는,

방법.

청구항 12

하드웨어와 결합되어 제1항 내지 제11항 중 어느 하나의 항의 방법을 실행시키기 위하여 컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램.

청구항 13

전자 장치에 있어서,

의료 데이터의 복수의 피처들(features) 중에서 트리 기반의 앙상블 모델(tree-based ensemble model)에 기초하여 선택된 피처를 가지는 부분 입력 데이터(partial input data)를 생성하고, 복수의 부분 입력 데이터들을 포함하는 입력 데이터(input data)를 시퀀스 데이터(sequence data)로 생성하며, 상기 생성된 시퀀스 데이터에 시퀀스 모델(sequence model)을 적용함으로써 출력 데이터(output data)를 산출하는 프로세서

를 포함하는 전자 장치.

청구항 14

제13항에 있어서,

상기 프로세서는,

상기 의료 데이터에 상기 트리 기반의 앙상블 모델을 적용하여 산출된 임시 출력 데이터와 참 값을 비교함으로

써 상기 트리 기반의 앙상블 모델을 트레이닝시키고,

상기 트레이닝된 트리 기반의 앙상블 모델에 대한 상기 복수의 피처들 각각의 피처 중요도에 기초하여, 상기 복수의 피처들 중 적어도 하나의 피처를 선택하는,

전자 장치.

청구항 15

제13항에 있어서,

상기 프로세서는,

상기 의료 데이터의 복수의 피처들 중 선택된 피처를 가지는 임시 부분 입력 데이터에 상기 트리 기반의 앙상블 모델을 적용하여 산출된 임시 출력 데이터와 참 값을 비교함으로써 상기 트리 기반의 앙상블 모델을 트레이닝시키고,

상기 임시 부분 입력 데이터의 각 피처의 상기 트레이닝된 트리 기반의 앙상블 모델에 대한 피처 중요도에 기초하여, 상기 임시 부분 입력 데이터의 피처 중 적어도 하나의 피처를 선택하며,

상기 선택된 적어도 하나의 피처를 가지는 데이터로 상기 임시 부분 입력 데이터를 업데이트하고,

상기 업데이트된 임시 부분 입력 데이터에 적용됨으로써 임시 출력 데이터를 산출하는 앙상블 모델로 상기 트리 기반의 앙상블 모델을 업데이트하는,

전자 장치.

청구항 16

제13항에 있어서,

상기 프로세서는,

부분 입력 데이터의 피처들 간의 시간 순서와 독립적으로 임시 출력 데이터를 산출하는 상기 트리 기반의 앙상블 모델에 기초하여 피처를 선택하는,

전자 장치.

청구항 17

제13항에 있어서,

상기 프로세서는,

상기 복수의 부분 입력 데이터들이 대응하는 시간대들의 순서에 따라 복수의 부분 입력 데이터들을 결합(combine)함으로써 상기 입력 데이터를 생성하는,

전자 장치.

청구항 18

제13항에 있어서,

상기 프로세서는,

상기 생성된 입력 데이터에 포함된 상기 복수의 부분 입력 데이터들의 결합 순서에 따라 다른 출력 데이터를 산출 가능한 상기 시퀀스 모델을 이용하여 상기 출력 데이터를 산출하는,

전자 장치.

청구항 19

제13항에 있어서,

상기 프로세서는,

상기 트리 기반의 앙상블 모델로서 XGBoost, AdaBoost, GBM(Gradient Boosting Machine), LightGBM, CatBoost, 또는 histogram-based gradient boost 중 적어도 하나를 포함하는 부스팅 알고리즘(boosting algorithm) 기반의 모델에 기초하여 상기 부분 입력 데이터를 생성하는,

전자 장치.

청구항 20

제13항에 있어서,

상기 프로세서는,

상기 시퀀스 모델로서, RNN(Recurrent Neural Network), LSTM(Long Short Term Memory), S2S(Sequence to Sequence model), GRU(Gated Recurrent Unit), 또는 Transformer 중 적어도 하나를 포함하는 기계 학습 모델을 상기 입력 데이터에 적용함으로써 상기 출력 데이터를 산출하는,

전자 장치.

발명의 설명

기 술 분 야

[0001] 이하, 트리 기반의 앙상블 모델을 이용하여 시퀀스 모델의 입력 데이터의 피처들을 선택하는 것에 대하여, 보다 구체적으로는, 입력 데이터의 복수의 부분 입력 데이터들 각각의 피처를 선택하는 것에 대하여, 개시한다.

배 경 기 술

[0003] 신경망 모델(Neural Network model)은 생물학에서의 뉴런 구조로부터 착안된 지도 학습(supervised learning) 알고리즘이다. 신경망 모델의 기본적인 작동 원리는 여러 개의 뉴런들을 상호 연결하여 입력값에 대한 최적의 출력값을 예측하는 것이다. 통계적인 관점에서 보면 신경망 모델은 입력 변수의 선형 결합에 비선형 함수를 취하는 사영추적회귀로 볼 수 있다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0005] 일 실시예에 따른 전자 장치에 의하여 수행되는 방법은, 의료 데이터의 복수의 피처들(features) 중에서 트리 기반의 앙상블 모델(tree-based ensemble model)에 기초하여 부분 입력 데이터(partial input data)의 피처를 선택하는 단계; 상기 선택된 피처를 가지는 복수의 부분 입력 데이터들을 포함하는 입력 데이터(input data)를 시퀀스 데이터(sequence data)로 생성하는 단계; 및 상기 생성된 입력 데이터에 시퀀스 모델(sequence model)을

적용함으로써 출력 데이터(output data)를 산출하는 단계를 포함할 수 있다.

[0006] 상기 부분 입력 데이터의 피처를 선택하는 단계는, 상기 의료 데이터에 상기 트리 기반의 앙상블 모델을 적용하여 산출된 임시 출력 데이터와 참 값(ground truth)을 비교함으로써 상기 트리 기반의 앙상블 모델을 트레이닝시키는 단계; 및 상기 트레이닝된 트리 기반의 앙상블 모델에 대한 상기 복수의 피처들 각각의 피처 중요도에 기초하여, 상기 복수의 피처들 중 적어도 하나의 피처를 선택하는 단계를 포함할 수 있다.

[0007] 상기 부분 입력 데이터의 피처를 선택하는 단계는, 트레이닝된 트리 기반의 앙상블 모델에 대한 피처 중요도에 기초하여, 상기 의료 데이터의 상기 복수의 피처들 중에서 미리 결정된 개수의 피처를 선택하는 단계를 포함할 수 있다.

[0008] 상기 부분 입력 데이터의 피처를 선택하는 단계는, 상기 의료 데이터의 복수의 피처들 중 선택된 피처를 가지는 임시 부분 입력 데이터에 상기 트리 기반의 앙상블 모델을 적용하여 산출된 임시 출력 데이터와 참 값을 비교함으로써 상기 트리 기반의 앙상블 모델을 트레이닝시키는 단계; 상기 트레이닝된 트리 기반의 앙상블 모델에 대하여 계산된 상기 임시 부분 입력 데이터의 각 피처의 피처 중요도(feature importance)에 기초하여, 상기 임시 부분 입력 데이터의 피처 중 적어도 하나의 피처를 선택하는 단계; 상기 선택된 적어도 하나의 피처를 가지는 데이터로 상기 임시 부분 입력 데이터를 업데이트하는 단계; 및 상기 업데이트된 임시 부분 입력 데이터에 적용됨으로써 임시 출력 데이터를 산출하는 앙상블 모델로 상기 트리 기반의 앙상블 모델을 업데이트하는 단계를 포함할 수 있다.

[0009] 상기 부분 입력 데이터의 피처를 선택하는 단계는, 부분 입력 데이터의 피처들 간의 시간 순서와 독립적으로 임시 출력 데이터를 산출하는 상기 트리 기반의 앙상블 모델에 기초하여 피처를 선택하는 단계를 포함할 수 있다.

[0010] 상기 입력 데이터를 생성하는 단계는, 상기 복수의 부분 입력 데이터들이 대응하는 시간대들의 시간 순서에 따라 복수의 부분 입력 데이터들을 결합(combine)함으로써 상기 입력 데이터를 생성하는 단계를 포함할 수 있다.

[0011] 상기 출력 데이터를 산출하는 단계는, 상기 생성된 입력 데이터에 포함된 상기 복수의 부분 입력 데이터들의 결합 순서에 따라 다른 출력 데이터를 산출 가능한 상기 시퀀스 모델을 이용하여 상기 출력 데이터를 산출하는 단계를 포함할 수 있다.

[0012] 상기 출력 데이터를 산출하는 단계는, 복수의 부분 입력 데이터들이 대응하는 시간대들 중 마지막 시간대의 종료 시각 이후의 출력 시간대에 대응하는 출력 데이터를 산출하는 단계를 포함할 수 있다.

[0013] 상기 부분 입력 데이터의 피처를 선택하는 단계는, 상기 트리 기반의 앙상블 모델로서 XGBoost, AdaBoost, GBM(Gradient Boosting Machine), LightGBM, CatBoost, 또는 histogram-based gradient boost 중 적어도 하나를 포함하는 부스팅 알고리즘(boosting algorithm) 기반의 모델에 기초하여 상기 부분 입력 데이터의 피처를 선택하는 단계를 포함할 수 있다.

[0014] 상기 출력 데이터를 산출하는 단계는, 상기 시퀀스 모델로서, RNN(Recurrent Neural Network), LSTM(Long Short Term Memory), ARIMA(AutoregRessive Integrated Moving Average), Seq2Seq(Sequence to Sequence model), GRU(Gated Recurrent Unit), 또는 Transformer 중 적어도 하나를 포함하는 기계 학습 모델을 상기 입력 데이터에 적용함으로써 상기 출력 데이터를 산출하는 단계를 포함할 수 있다.

[0015] 상기 출력 데이터를 산출하는 단계는, 추정 사구체 여과율(estimated Glomerular Filtration Rate; eGFR)을 포함하는 상기 출력 데이터를 산출하는 단계를 포함할 수 있다.

[0017] 일 실시예에 따른 전자 장치는, 의료 데이터의 복수의 피처들(features) 중에서 트리 기반의 앙상블 모델(tree-based ensemble model)에 기초하여 선택된 피처를 가지는 부분 입력 데이터(partial input data)를 생성하고, 복수의 부분 입력 데이터들을 포함하는 입력 데이터(input data)를 시퀀스 데이터(sequence data)로 생성하며, 상기 생성된 시퀀스 데이터에 시퀀스 모델(sequence model)을 적용함으로써 출력 데이터(output data)를 산출하는 프로세서를 포함할 수 있다.

[0018] 상기 프로세서는, 상기 의료 데이터에 상기 트리 기반의 앙상블 모델을 적용하여 산출된 임시 출력 데이터와 참 값을 비교함으로써 상기 트리 기반의 앙상블 모델을 트레이닝시키고, 상기 트레이닝된 트리 기반의 앙상블 모델에 대한 상기 복수의 피처들 각각의 피처 중요도에 기초하여, 상기 복수의 피처들 중 적어도 하나의 피처를 선택할 수 있다.

[0019] 상기 프로세서는, 상기 의료 데이터의 복수의 피처들 중 선택된 피처를 가지는 임시 부분 입력 데이터에 상기 트리 기반의 앙상블 모델을 적용하여 산출된 임시 출력 데이터와 참 값을 비교함으로써 상기 트리 기반의 앙상블 모델을 트레이닝시키고, 상기 임시 부분 입력 데이터의 각 피처의 상기 트레이닝된 트리 기반의 앙상블 모델에 대한 피처 중요도에 기초하여, 상기 임시 부분 입력 데이터의 피처 중 적어도 하나의 피처를 선택하며, 상기 선택된 적어도 하나의 피처를 가지는 데이터로 상기 임시 부분 입력 데이터를 업데이트하고, 상기 업데이트된 임시 부분 입력 데이터에 적용됨으로써 임시 출력 데이터를 산출하는 앙상블 모델로 상기 트리 기반의 앙상블 모델을 업데이트할 수 있다.

[0020] 상기 프로세서는, 부분 입력 데이터의 피처들 간의 시간 순서와 독립적으로 임시 출력 데이터를 산출하는 상기 트리 기반의 앙상블 모델에 기초하여 피처를 선택할 수 있다.

[0021] 상기 프로세서는, 상기 복수의 부분 입력 데이터들이 대응하는 시간대들의 순서에 따라 복수의 부분 입력 데이터들을 결합(combine)함으로써 상기 입력 데이터를 생성할 수 있다.

[0022] 상기 프로세서는, 상기 생성된 입력 데이터에 포함된 상기 복수의 부분 입력 데이터들의 결합 순서에 따라 다른 출력 데이터를 산출 가능한 상기 시퀀스 모델을 이용하여 상기 출력 데이터를 산출할 수 있다.

[0023] 상기 프로세서는, 상기 트리 기반의 앙상블 모델로서 XGBoost, AdaBoost, GBM(Gradient Boosting Machine), LightGBM, CatBoost, 또는 histogram-based gradient boost 중 적어도 하나를 포함하는 부스팅 알고리즘(boosting algorithm) 기반의 모델에 기초하여 상기 부분 입력 데이터를 생성할 수 있다.

[0024] 상기 프로세서는, 상기 시퀀스 모델로서, RNN(Recurrent Neural Network), LSTM(Long Short Term Memory), S2S(Sequence to Sequence model), GRU(Gated Recurrent Unit), 또는 Transformer 중 적어도 하나를 포함하는 기계 학습 모델을 상기 입력 데이터에 적용함으로써 상기 출력 데이터를 산출할 수 있다.

도면의 간단한 설명

[0026] 도 1은 다양한 실시예들에 따른 전자 장치의 출력 데이터의 산출 동작을 설명하기 위한 도면이다.

도 2는 다양한 실시예들에 따른 전자 장치의 출력 데이터의 산출 동작을 나타내는 흐름도다.

도 3은 다양한 실시예들에 따른 트리 기반의 앙상블 모델을 설명하기 위한 도면이다.

도 4은 다양한 실시예들에 따른 시퀀스 모델을 설명하기 위한 도면이다.

도 5는 다양한 실시예들에 따른 전자 장치의 부분 입력 데이터의 피처를 선택하는 동작을 설명하기 위한 도면이다.

도 6은 다양한 실시예들 및 비교 실시예에 따른 시퀀스 모델의 성능을 비교하기 위한 도면이다.

도 7은 다양한 실시예들 및 비교 실시예에 따른 시퀀스 모델의 성능을 비교하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0027] 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.

[0028] 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.

[0029] 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.

[0030] 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 존재함으로 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.

[0031] 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.

[0032] 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.

[0034] 도 1은 다양한 실시예들에 따른 전자 장치의 출력 데이터의 산출 동작을 설명하기 위한 도면이다.

[0035] 일 실시예에 따른 전자 장치(100)는, 프로세서(110) 및 메모리(120)를 포함할 수 있다.

[0036] 프로세서(110)는 의료 데이터(medical data)(111)의 복수의 피처들(medical features) 중에서 트리 기반의 앙상블 모델(tree-based ensemble model)(112)에 기초하여 부분 입력 데이터(partial input data)(114a, 114b, 114c)의 피처를 선택할 수 있다. 프로세서(110)는, 복수의 부분 입력 데이터들(114a, 114b, 114c)을 포함하는 입력 데이터(114)를 시퀀스 데이터(sequence data)로 생성할 수 있다. 프로세서(110)는, 입력 데이터(114)에 시퀀스 모델(sequence model) (115)을 적용함으로써 출력 데이터(output data)(116)를 산출할 수 있다.

[0037] 의료 데이터는, 복수의 피처들을 가지는 데이터로서, 예시적으로 전자의무기록(Electronic Medical Record; EMR)을 포함할 수 있다. 피처(예: 의료 피처)는 대상자(예: 환자)의 의료 상황과 관련된 정보를 분류한 개별 항목 및/또는 카테고리를 나타낼 수 있다. 의료 피처는, 예를 들어, 수술 피처(operation feature), 처치 피처(procedure feature), 의료영상전송시스템 피처(Picture Archiving and Communication System feature; PACS feature), 진단 피처(diagnosis feature), 복약 피처(medication feature), 검사 피처(laboratory feature), 신체 정보 피처(physical information feature), 또는 중환자실 체류기간 피처(length of stay of intensive care unit feature; LOS of ICU feature) 중 하나 또는 둘 이상의 조합을 포함할 수 있다.

[0038] 메모리(120)는 의료 데이터(111) 및/또는 의료 데이터(111)의 복수의 피처들을 저장할 수 있다. 메모리(120)는 트리 기반의 앙상블 모델(112) 및 시퀀스 모델(115)의 구조 및/또는 파라미터들을 저장할 수 있다. 메모리(120)는 임시 출력 데이터(113)에 관한 정보, 부분 입력 데이터(114a, 114b, 114c)에 관한 정보, 입력 데이터(114)에 관한 정보, 및/또는 출력 데이터(116)에 관한 정보를 저장할 수 있다. 또한, 메모리(120)는 의료 데이터의 처리를 통해 출력 데이터를 산출하기 위한 방법을 수행하기 위해 요구되는 데이터 및/또는 정보를 일시적으로 및/또는 영구적으로 저장할 수 있다.

[0040] 도 2는 다양한 실시예들에 따른 전자 장치의 출력 데이터의 산출 동작을 나타내는 흐름도다.

[0041] 단계(210)에서, 프로세서는 트리 기반의 앙상블 모델에 기초하여 부분 입력 데이터의 피처를 선택할 수 있다. 프로세서는, 의료 데이터의 복수의 피처들 중에서 부분 입력 데이터의 피처를 선택할 수 있다. 후술하겠으나, 트리 기반의 앙상블 모델은 부분 입력 데이터를 입력받을 수 있고, 복수의 부분 입력 데이터들이 결합됨으로써 시퀀스 모델의 입력 데이터가 생성될 수 있다.

[0042] 앙상블 모델은, 복수의 약한 분류기들(week classifiers)을 이용하는 모델을 의미할 수 있다. 트리 기반의 앙상블 모델은, 복수의 트리들을 생성하고, 생성된 복수의 트리들에 기초하여 산출된 출력들을 이용하는 모델을 의미할 수 있다. 일 실시예에 따른 전자 장치는, 앙상블 모델을 이용함으로써, 하나의 약한 분류기를 이용하는 경우에 발생 가능한 오버피팅(overfitting)을 감소시킬 수 있다. 트리 기반의 앙상블 모델은 도 3에서 후술한다.

[0043] 일 실시예에 따르면, 프로세서는 의료 데이터에 적용됨으로써 임시 출력 데이터를 산출하는데 이용되는 트리 기반의 앙상블 모델을 트레이닝시킬 수 있다. 예를 들어, 프로세서는 의료 데이터에 트리 기반의 앙상블 모델을 적용하여 임시 출력 데이터를 산출할 수 있다. 프로세서는 임시 출력 데이터와 참 값(ground truth)을 비교함으로써 트리 기반의 앙상블 모델을 트레이닝시킬 수 있다. 프로세서는, 임시 출력 데이터와 참 값을 비교하여 산출된 목적 함수에 기초하여 트리 기반의 앙상블 모델의 파라미터를 업데이트할 수 있다.

[0044] 일 실시예에 따르면, 프로세서는, 피처 중요도(feature importance)에 기초하여 복수의 피처들 중 부분 입력 데이터의 피처를 선택할 수 있다. 프로세서는, 트레이닝된 트리 기반의 앙상블 모델에 대한, 의료 데이터의 복수의 피처들 각각의 피처 중요도(feature importance)를 계산할 수 있다. 예시적으로, 의료 데이터가 제1 피처 및 제2 피처를 가지는 경우, 프로세서는 트레이닝된 트리 기반의 앙상블 모델에 대한 제1 피처의 피처 중요도를 계산하고, 트레이닝된 트리 기반의 앙상블 모델에 대한 제2 피처의 피처 중요도를 계산할 수 있다.

[0045] 예를 들어, 프로세서는 계산된 피처 중요도에 기초하여 복수의 피처들 중 적어도 하나의 피처를 선택할 수 있다. 프로세서는, 의료 데이터의 복수의 피처들 중에서 미리 결정된 개수의 피처를 선택할 수 있다. 프로세서는, 복수의 피처들 중 피처 중요도가 큰 순서대로 미리 결정된 개수의 피처를 선택할 수 있다. 복수의 피처들 중 부분 입력 데이터의 피처로 선택된 피처들은, 나머지 피처들보다 큰 피처 중요도를 가질 수 있다.

[0046] 예를 들어, 프로세서는, 계산된 피처 중요도에 기초하여 복수의 피처들 중 임시 부분 입력 데이터의 피처를 선택함으로써 임시 부분 입력 데이터 및 트리 기반의 앙상블 모델을 업데이트할 수 있다. 임시 부분 입력 데이터 및 트리 기반의 앙상블 모델의 업데이트를 통한 부분 입력 데이터의 피처를 선택하는 동작은 도 5에서 후술한다.

[0047] 후술하겠으나, 시퀀스 모델은 입력 데이터의 시계열 요인에 기초하여 출력 데이터를 산출할 수 있다. 시퀀스 모델의 입력 데이터 및/또는 부분 입력 데이터의 피처를 선택하기 위하여 시퀀스 모델에 대한 피처 중요도를 이용하는 것 보다, 일 실시예에 따른 전자 장치는 트리 기반의 앙상블 모델에 대한 피처 중요도를 이용함으로써, 피처 중요도의 계산 및 피처의 선택을 보다 더 간단하게 수행 가능할 수 있다.

[0048] 단계(220)에서, 프로세서는 복수의 부분 입력 데이터들을 포함하는 입력 데이터를 시퀀스 데이터로 생성할 수 있다. 복수의 부분 입력 데이터들 각각은, 부분 입력 데이터의 피처로서 선택된 피처를 가질 수 있다. 부분 입력 데이터는, 의료 데이터의 적어도 일부로서 대상자(예: 환자)의 의료 상황과 관련된 정보를 포함할 수 있다. 일 실시예에 따르면, 부분 입력 데이터는 시간대에 대응할 수 있다.

[0049] 부분 입력 데이터는, 대응하는 시간대(예: 부분 입력 시간대)에서의 대상자의 의료 상황과 관련된 정보를 포함할 수 있다. 예시적으로, 부분 입력 데이터는, 부분 입력 시간대에 포함된 시각에서 측정 및/또는 기록된 대상자의 의료 상황과 관련된 정보(예: 체중, 혈당, 등)를 포함할 수 있다. 예시적으로, 부분 입력 데이터는, 부분 입력 시간대에 걸친 평균적인 대상자의 의료 정보(예: 평균 심박수, 평균 혈압 등)를 포함할 수 있다. 부분 입력 시간대는, 미리 결정된 시간 길이를 가질 수 있다. 예시적으로, 부분 입력 시간대는 24시간의 시간 길이를 가질 수 있으나, 이에 한정하는 것은 아니고, 설계에 따라 12시간, 36시간, 48시간 등 다른 시간 길이를 가지는 것으로 변경될 수 있다.

[0050] 시퀀스 데이터는, 시간 순서를 가지는 데이터로서, 시간의 흐름에 따라 측정 및/또는 기록된 데이터를 의미할 수 있다. 시퀀스 데이터는, 복수의 서브 데이터들(예: 부분 입력 데이터들)을 포함할 수 있다. 시퀀스 데이터는, 복수의 서브 데이터들 각각의 측정 및/또는 기록된 시각 또는 시간대에 기초하여, 복수의 서브 데이터들을 시간 순서대로 포함할 수 있다.

[0051] 일 실시예에 따르면, 프로세서는 복수의 부분 입력 데이터들을 시간 순서대로 결합함으로써 입력 데이터를 시퀀스 데이터로 생성할 수 있다. 복수의 부분 입력 데이터들은, 연속된 시간대들에 대응할 수 있다. 프로세서는, 복수의 부분 입력 데이터들이 대응하는 시간대들의 순서에 따라 복수의 부분 입력 데이터들을 결합함으로써, 입력 데이터를 생성할 수 있다. 예시적으로, 제1 부분 입력 데이터는 제1 시간대에 대응하고, 제2 부분 입력 데이터는 제1 시간대 이후의 제2 시간대에 대응할 수 있다. 제1 시간대는 제2 시간대와 연속할 수 있다. 다시 말해, 제1 시간대의 종료 시각은 제2 시간대의 시작 시각과 같을 수 있다. 프로세서는, 제1 시간대가 제2 시간대보다 선행하는(preceding) 것에 기초하여, 제1 부분 입력 데이터 이후에 제2 부분 입력 데이터를 결합함으로써 입력 데이터를 생성할 수 있다.

[0052] 단계(230)에서, 프로세서는 입력 데이터에 시퀀스 모델을 적용함으로써 출력 데이터를 산출할 수 있다. 시퀀스 모델은, 입력 시간대(예: 과거 시간대 및/또는 현재 시간대)에 대응하는 데이터에 적용됨으로써 출력 시간대(예: 미래 시간대) 이후의 출력 시간대에 대응하는 데이터를 예측에 이용될 수 있다. 예를 들어, 프로세서는 입력 데이터에 시퀀스 모델을 적용함으로써 출력 데이터를 산출함으로써, 출력 시간대에 대응하는 데이터를 예측할 수 있다. 시퀀스 모델은 도 4에서 후술한다.

[0053] 입력 데이터는, 입력 시간대에 대응할 수 있다. 입력 데이터는 복수의 시간대들에 대응하는 부분 입력 데이터들을 포함할 수 있고, 복수의 시간대들은 연속할 수 있다. 입력 시간대의 시작 시각은, 복수의 시간대들 중 첫

번째 시간대의 시작 시각으로 결정될 수 있다. 입력 시간대의 종료 시각은, 복수의 시간대들 중 마지막 시간대의 종료 시각으로 결정될 수 있다. 결과적으로, 입력 데이터는, 포함된 복수의 부분 입력 데이터들이 대응하는 시간대들의 합집합인 입력 시간대에 대응할 수 있다.

[0054] 출력 데이터는, 대응하는 출력 시간대에 측정 및/또는 기록될 것으로 예측되는 대상자(예: 환자)의 의료 상황과 관련된 값을 포함하는 데이터를 나타낼 수 있다. 출력 시간대는 입력 시간대 이후의 시간대일 수 있다. 예를 들어, 출력 시간대는, 입력 시간대의 복수의 부분 입력 시간대들 중 마지막 부분 입력 시간대의 종료 시각 이후의 시간대일 수 있다. 일 실시예에 따르면, 프로세서는 추정 사구체 여과율(estimated Glomerular Filtration Rate; eGFR)을 포함하는 출력 데이터를 산출할 수 있다. 다만, 출력 데이터를 이에 한정하는 것은 아니고, 출력 데이터는 진단 검사(laboratory test)에 따라 시간대 별로 수집 가능한 의료 상황과 관련된 값을 포함할 수 있다.

[0055] 출력 시간대의 시작 시각은 입력 시간대에 기초하여 결정될 수 있다. 예를 들어, 출력 시간대의 시작 시각은 입력 시간대의 종료 시각과 같을 수 있다. 다른 예를 들어, 출력 시간대의 시작 시각은, 입력 시간대의 종료 시각으로부터 미리 결정된 시간 길이 이후의 시각일 수 있다.

[0056] 출력 시간대의 종료 시각은, 출력 시간대의 시작 시각 및 출력 시간대의 시간 길이에 기초하여 결정될 수 있다. 예를 들어, 출력 시간대의 종료 시각은, 출력 시간대의 시작 시각으로부터 미리 결정된 시간 길이 이후의 시각일 수 있다. 출력 시간대의 시간 길이는, 다시 말해, 출력 시간대의 시작 시각으로부터 출력 시간대의 종료 시각까지의 시간 길이는, 부분 입력 데이터가 대응하는 시간대(예: 부분 입력 시간대)의 길이(예: 24시간)와 같을 수 있다. 다만, 출력 시간대의 시간 길이가 이에 한정되는 것은 아니고, 부분 입력 시간대의 시간 길이와 독립적인 시간 길이로 설계에 따라 변경될 수 있다.

[0058] 도 3은 다양한 실시예들에 따른 트리 기반의 앙상블 모델을 설명하기 위한 도면이다.

[0059] 일 실시예에 따르면, 트리 기반의 앙상블 모델(320)은, 부분 입력 데이터(310)에 적용됨으로써 임시 출력 데이터(330)를 산출하는 모델을 나타낼 수 있다. 트리 기반의 앙상블 모델(320)은, 입력 데이터의 시계열 요인(time series factor)에 기초하여 출력 데이터를 산출하는 시퀀스 모델과 달리, 부분 입력 데이터(310)의 피처들 간의 순서(예: 시간 순서, 결합 순서)와 독립적으로 임시 출력 데이터(330)를 산출하는 모델을 포함할 수 있다. 예를 들어, 부분 입력 데이터(310)는 복수의 피처들(예: 제1 피처(311), 제2 피처(312), 제3 피처(313), 제4 피처(314), 제5 피처(315), 및 제6 피처(316))을 가질 수 있다. 트리 기반의 앙상블 모델(320)은, 부분 입력 데이터(310)의 피처들의 측정 및/또는 기록된 시간 순서와 독립적으로 임시 출력 데이터(330)를 산출할 수 있다.

[0060] 이하, 설명의 편의를 위하여, 부분 입력 데이터(310) 및 해당 부분 입력 데이터(310)에 트리 기반의 앙상블 모델(320)을 적용함으로써 산출된 임시 출력 데이터(330)를 데이터 쌍(data pair)으로 페어링(pairing)하여 설명한다.

[0061] 도 2에서 전술한 바와 같이, 부분 입력 데이터(310)는, 의료 데이터의 적어도 일부로서, 대응하는 시간대(예: 부분 입력 시간대(341))에서의 대상자의 의료 상황과 관련된 정보를 포함할 수 있다. 복수의 부분 입력 데이터(310)들이 결합됨으로써, 시퀀스 모델의 입력 데이터가 생성될 수 있다. 부분 입력 데이터(310)는 미리 결정된 시간 길이를 가지는 부분 입력 시간대(341)에 대응할 수 있다.

[0062] 임시 출력 데이터(330)는, 임시 출력 데이터(330)가 대응하는 시간대(342)에 측정 및/또는 기록될 것으로 예측되는 대상자(예: 환자)의 의료 상황과 관련된 값을 포함하는 데이터를 나타낼 수 있다. 임시 출력 데이터(330)는 페어링된 부분 입력 데이터(310)가 대응하는 부분 입력 시간대(341) 이후의 시간대에 대응할 수 있다.

[0063] 예를 들어, 임시 출력 데이터(330)가 대응하는 시간대(342)의 시작 시각은, 페어링된 부분 입력 데이터(310)가 대응하는 부분 입력 시간대(341)의 종료 시각과 같을 수 있다. 임시 출력 데이터(330)가 대응하는 시간대(342)의 종료 시각은, 임시 출력 데이터(330)가 대응하는 시간대(342)의 시작 시각으로부터 미리 결정된 시간 길이 이후의 시각일 수 있다. 임시 출력 데이터(330)가 대응하는 시간대(342)의 시간 길이는, 다시 말해, 임시 출력 데이터(330)가 대응하는 시간대(342)의 시작 시각으로부터 종료 시각까지의 시간 길이는, 부분 입력 시간대(341)의 길이와 같을 수 있다. 다만, 임시 출력 데이터(330)가 대응하는 시간대(342)의 시간 길이가 이에 한정되는 것은 아니고, 부분 입력 시간대(341)의 시간 길이와 독립적인 시간 길이로 설계에 따라 변경될 수 있다.

[0064] 일 실시예에 따르면, 프로세서는, 트리 기반의 앙상블 모델(320)로서 부스팅 알고리즘(boosting algorithm) 기반의 모델에 기초하여 부분 입력 데이터(310)의 피처를 선택할 수 있다.

[0065] 부스팅 알고리즘은, 복수의 약한 분류기들 중 선행하는 분류기가 잘못 예측한 결과에 기초하여 후속하는 분류기의 가중치를 변경함으로써, 선행하는 분류기에 의하여 잘못 분류된 데이터에 집중하여 새로운 분류 규칙을 만드는 특징을 가질 수 있다. 예를 들어, 프로세서는 복수개의 분류기들을 순차적으로 트레이닝시킬 수 있다. 프로세서는, 선행하는 분류기의 출력 중 예측(prediction)과 참 값이 다른 경우, 후속하는 분류기는 참 값을 출력할 수 있도록 후속하는 분류기의 가중치를 조정할 수 있다.

[0066] 예시적으로, 부스팅 알고리즘 기반의 모델은, XGBoost, AdaBoost, GBM(Gradient Boosting Machine), LightGBM, CatBoost, 또는 histogram-based gradient boost 중 적어도 하나를 포함할 수 있다.

[0067] 일 실시예에 따르면, 프로세서는 트리 기반의 앙상블 모델(320)로서 배깅 알고리즘(bagging algorithm) 기반의 모델에 기초하여 부분 입력 데이터(310)의 피처를 선택할 수 있다.

[0068] 배깅 알고리즘은, 원본 데이터를 랜덤 샘플링함으로써 생성된 복수의 샘플 데이터들 각각에 기초하여 복수의 약한 분류기들(예: 트리들)을 병렬적으로 트레이닝하는 동작을 포함할 수 있다. 배깅 알고리즘 기반의 모델은, 복수의 약한 분류기들 각각의 결과를 결합함으로써 결과를 출력할 수 있다. 배깅 알고리즘 기반의 모델은, 복수의 결과들의 평균(average) 및/또는 다중 투표(majority vote)을 통해 복수의 결과들을 결합할 수 있다.

[0069] 예시적으로, 배깅 알고리즘 기반의 모델은, 랜덤 포레스트(random forest)를 포함할 수 있다.

[0071] 도 4은 다양한 실시예들에 따른 시퀀스 모델을 설명하기 위한 도면이다.

[0072] 일 실시예에 따르면, 시퀀스 모델(420)은 입력 데이터(410)에 적용됨으로써 출력 데이터(430)를 산출하는 모델을 나타낼 수 있다. 입력 데이터(410)는, 대응하는 입력 시간대에 측정 및/또는 기록된 대상자의 의료 상황과 관련된 값을 포함하는 데이터를 나타낼 수 있다. 출력 데이터(430)는, 대응하는 출력 시간대에 측정 및/또는 기록될 것으로 예측되는 대상자의 의료 상황과 관련된 값을 포함하는 데이터를 나타낼 수 있다. 출력 시간대(예: 출력 시간대(450a), 출력 시간대(450b), 출력 시간대(450c))는 입력 시간대(440) 이후의 시간대로서, 시퀀스 모델(420)의 설계에 따라 결정될 수 있다.

[0073] 일 실시예에 따르면, 입력 데이터(410)는, 복수의 부분 입력 데이터들(예: 제1 부분 입력 데이터(411), 제2 부분 입력 데이터(412), 제3 부분 입력 데이터(413))을 포함할 수 있다. 입력 데이터(410)는, 복수의 부분 입력 데이터들을 시간 순서에 따라 결합된 시퀀스 데이터를 포함할 수 있다. 일 실시예에 따르면, 프로세서는, 제1 부분 입력 데이터(411), 제2 부분 입력 데이터(412), 및 제3 부분 입력 데이터(413)를 시간 순서에 따라 결합함으로써 입력 데이터(410)를 생성할 수 있다. 복수의 부분 입력 데이터들에 대한 시간 순서는, 부분 입력 데이터(410)가 대응하는 시간대 간의 시간 순서에 의하여 결정될 수 있다.

[0074] 예시적으로, 도 4에서 나타난 바와 같이, 제1 부분 입력 데이터(411)는 제1 부분 입력 시간대(441)에 대응할 수 있다. 제2 부분 입력 데이터(412)는 제2 부분 입력 시간대(442)에 대응할 수 있다. 제3 부분 입력 데이터(413)는 제3 부분 입력 시간대(443)에 대응할 수 있다. 프로세서는, 복수의 부분 입력 데이터들을, 복수의 부분 입력 시간대들의 시간 순서에 따라, 제1 부분 입력 데이터(411), 제2 부분 입력 데이터(412), 및 제3 부분 입력 데이터(413)의 순서로 결합함으로써 입력 데이터(410)를 생성할 수 있다. 생성된 입력 데이터(410)는, 복수의 부분 입력 시간대들(예: 제1 부분 입력 시간대(441), 제2 부분 입력 시간대(442), 및 제3 부분 입력 시간대(443))을 포함하는 입력 시간대(440)에 대응할 수 있다.

[0075] 프로세서는, 입력 데이터(410)에 시퀀스 모델(420)을 적용함으로써, 입력 시간대(440) 이후의 출력 시간대에 대응하는 출력 데이터(430)를 산출할 수 있다.

[0076] 일 실시예에 따르면, 출력 시간대(450a)의 시작 시각은 입력 시간대(440)의 종료 시각일 수 있다. 출력 시간대(450a)의 종료 시각은, 출력 시간대(450a)의 시작 시각으로부터 미리 결정된 시간 길이 이후의 시각일 수 있다. 다시 말해, 출력 시간대(450a)는 입력 시간대의 종료 시각으로부터 미리 결정된 시간 길이 이후의 시각까지의 시간대일 수 있다. 시퀀스 모델(420)은, 입력 시간대(440)에 대응하는 입력 데이터(410)에 적용되는 경우, 입력 시간대(440)의 종료 시각으로부터 미리 결정된 시간 길이 이후의 시각까지의 출력 시간대(450a)에 대응하는

출력 데이터(430)를 산출하도록 트레이닝될 수 있다. 프로세서는, 시퀀스 모델(420)을 이용하여 입력 시간대(440)에 대응하는 입력 데이터(410)에 기초하여 출력 시간대(450a)에 대응하는 출력 데이터(430)를 산출할 수 있다.

[0077] 일 실시예에 따르면, 출력 시간대의 시작 시각은 입력 시간대(440)의 종료 시각으로부터 제1 시간 길이 이후의 시각일 수 있다. 출력 시간대의 종료 시각은, 출력 시간대(450b)의 시작 시각으로부터 제2 시간 길이 이후의 시각일 수 있다.

[0078] 예를 들어, 제1 시간 길이 및 제2 시간 길이가 24시간인 경우, 시퀀스 모델(420)은 입력 시간대(440)에 대응하는 입력 데이터(410)에 적용되는 경우, 출력 시간대(450b)에 대응하는 출력 데이터(430)를 산출하도록 트레이닝될 수 있다. 프로세서는, 시퀀스 모델(420)을 이용하여 입력 시간대(440)에 대응하는 입력 데이터(410)에 기초하여 출력 시간대(450b)에 대응하는 출력 데이터(430)를 산출할 수 있다.

[0079] 예를 들어, 제1 시간 길이가 48시간이고 제2 시간 길이가 24시간인 경우, 시퀀스 모델(420)은 입력 시간대(440)에 대응하는 입력 데이터(410)에 적용되는 경우, 출력 시간대(450c)에 대응하는 출력 데이터 (430)를 산출하도록 트레이닝될 수 있다. 프로세서는, 시퀀스 모델(420)을 이용하여 입력 시간대(440)에 대응하는 입력 데이터(410)에 기초하여 출력 시간대(450c)에 대응하는 출력 데이터(430)를 산출할 수 있다.

[0080] 일 실시예에 따르면, 시퀀스 모델(420)은, 트리 기반의 앙상블 모델(예: 도 3의 트리 기반의 앙상블 모델(320))과 달리, 입력 데이터(410)의 시계열 요인에 기초하여 출력 데이터(430)를 산출하는 모델을 포함할 수 있다. 예를 들어, 입력 데이터(410)의 시계열 요인은, 입력 데이터(410)의 부분 입력 데이터들 간의 결합 순서를 포함할 수 있다. 다시 말해, 시퀀스 모델(420)은, 입력 데이터(410) 중에서, 부분 입력 데이터들뿐만 아니라, 부분 입력 데이터들의 결합 순서(예: 시간 순서, 무작위로 정렬된 다른 순서)에도 기초하여 출력 데이터(430)를 산출할 수 있다.

[0081] 일 실시예에 따르면, 시퀀스 모델(420)은, 입력 데이터(410)에 포함된 복수의 부분 입력 데이터들의 결합 순서에 따라 다른 출력 데이터를 산출 가능할 수 있다. 예를 들어, 시퀀스 모델(420)은, 동일한 부분 입력 데이터들을 서로 다른 순서로 결합함으로써 생성된 제1 입력 데이터 및 제2 입력 데이터에 각각 적용됨으로써, 서로 다른 출력 데이터를 산출할 수 있다. 제1 입력 데이터는, 부분 입력 데이터들을 시간 순서대로 결합함으로써 생성될 수 있다. 제1 입력 데이터는, 제1 부분 입력 데이터(411), 제2 부분 입력 데이터(412), 및 제3 부분 입력 데이터(413)의 순서로 결합됨으로써 생성될 수 있다. 제2 입력 데이터는, 시간 순서와 독립적으로 결정된 순서로 결합됨으로써 생성될 수 있다. 예시적으로, 제2 입력 데이터는, 제2 부분 입력 데이터(412), 제3 부분 입력 데이터(413), 및 제1 부분 입력 데이터(411)의 순서로 결합됨으로써 생성될 수 있다. 프로세서는, 제1 입력 데이터에 시퀀스 모델(420)을 적용함으로써 제1 출력 데이터를 출력할 수 있다. 프로세서는, 제2 입력 데이터에 시퀀스 모델(420)을 적용함으로써 제1 출력 데이터와 다른 제2 출력 데이터를 출력할 수 있다. 시퀀스 모델(420)은, 제1 입력 데이터 및 제2 입력 데이터가 동일한 부분 입력 데이터들을 결합함으로써 생성된 입력 데이터들임에도 불구하고, 부분 입력 데이터들의 결합 순서에 차이가 있으므로, 서로 다른 복수의 출력 데이터들(예: 제1 출력 데이터 및 제2 출력 데이터)을 산출 가능할 수 있다.

[0082] 예시적으로, 시퀀스 모델(420)은, RNN(Recurrent Neural Network), LSTM(Long Short Term Memory), ARIMA(AutoregRessive Integrated Moving Average), Seq2Seq(Sequence to Sequence model), GRU(Gated Recurrent Unit), 또는 Transformer 중 적어도 하나를 포함할 수 있다.

[0084] 도 5는 다양한 실시예들에 따른 전자 장치의 부분 입력 데이터의 피처를 선택하는 동작을 설명하기 위한 도면이다.

[0085] 단계(510)에서, 프로세서는 트리 기반의 앙상블 모델을 트레이닝시킬 수 있다. 프로세서는 임시 부분 입력 데이터에 트리 기반의 앙상블 모델을 적용하여 산출된 임시 출력 데이터와 참 값을 비교할 수 있다. 프로세서는 임시 출력 데이터와 임시 부분 입력 데이터에 매핑된 참 값의 비교에 기초하여 목적 함수를 산출할 수 있다. 프로세서는, 산출된 목적 함수에 기초하여 트리 기반의 앙상블 모델의 파라미터를 업데이트할 수 있다.

[0086] 임시 부분 입력 데이터는, 의료 데이터의 적어도 일부로서, 부분 입력 데이터를 획득하기 위하여 이용될 수 있다. 예시적으로, 프로세서는, 초기에 임시 부분 입력 데이터를 의료 데이터의 복수의 피처들을 모두 가지는 것으로 설정할 수 있다.

[0087] 단계(520)에서, 프로세서는 피처 중요도(feature importance)에 기초하여, 임시 부분 입력 데이터의 피처 중 적어도 하나의 피처를 선택할 수 있다. 프로세서는 트레이닝된 트리 기반의 앙상블 모델에 대하여, 임시 부분 입력 데이터의 각 피처에 대한 피처 중요도를 계산할 수 있다. 프로세서는, 피처 중요도에 기초하여 임시 부분 입력 데이터의 피처들 중 미리 결정된 개수의 피처를 선택할 수 있다. 임시 부분 입력 데이터의 피처들 중에서, 선택된 미리 결정된 개수의 피처는, 나머지 피처보다 큰 피처 중요도를 가질 수 있다.

[0088] 피처 중요도는, 트레이닝된 트리 기반의 앙상블 모델의 출력 데이터 산출에 대한 임시 부분 입력 데이터의 각 피처의 영향력을 나타내는 값을 포함할 수 있다. 예시적으로, 피처 중요도는, 게인(gain), 커버(cover), 또는 웨이트(weight) 중 하나 또는 둘 이상의 조합을 포함할 수 있다.

[0089] 게인(gain)은, 각 피처가 트리 기반의 앙상블 모델 예측에 미친 영향력을 나타내는 값을 포함할 수 있다. 예를 들어, 대상 피처의 게인은 데이터에서 대상 피처의 값을 임의의 값으로 치환하면 원본 데이터보다 임시 기계 학습 모델의 예측 에러가 증가하는 정도를 포함할 수 있다. 대상 노드에 대한 대상 피처의 게인은, 대상 노드에서 대상 피처로 분기되지 않는 경우에 비하여, 트리의 대상 노드가 대상 피처로 분기되는 경우에 가지는 트리 기반의 앙상블 모델의 성능의 증가분을 포함할 수 있다. 평균 게인은, 트리 기반의 앙상블 모델에 포함된 복수의 노드들에 대한 게인들의 평균으로 계산될 수 있다. 총 게인(total gain)은, 트리 기반의 앙상블 모델에 포함된 복수의 노드들에 대한 게인들의 합(summation)으로 계산될 수 있다.

[0090] 커버(cover)는, 각 피처와 관련된 샘플의 상대적인 개수를 나타내는 값을 포함할 수 있다. 예시적으로, 입력 데이터는 4개의 피처들을 가질 수 있다. 트리 기반의 앙상블 모델의 입력 데이터 세트는 100개의 입력 데이터들을 가질 수 있다. 트리 기반의 앙상블 모델은, 제1 트리, 제2 트리, 및 제3 트리를 포함할 수 있다. 대상 피처의 커버는, 입력 데이터 세트에 포함된 100개의 입력 데이터들 중에서, 대상 피처에 기초하여 제1 트리에서 10개의 입력 데이터들이 분류되고 제2 트리에서 5개의 입력 데이터들이 분류되며 제3 트리에서 2개의 입력 데이터들이 분류되는 경우, 17로 계산될 수 있다. 대상 피처의 평균 커버는, 앙상블 모델에 포함된 트리들에 대한 대상 피처의 커버들의 평균으로 계산될 수 있다. 총 커버(total cover)는, 트리 기반의 앙상블 모델에 대한 복수의 피처들의 커버들의 합으로 계산될 수 있다.

[0091] 웨이트(weight)는, 트리 기반의 앙상블 모델의 구조에서, 각 피처에 기초한 노드 분기의 개수를 나타내는 값을 포함할 수 있다. 예시적으로, 트리 기반의 앙상블 모델은, 제1 트리, 제2 트리, 및 제3 트리를 포함할 수 있다. 대상 피처의 웨이트는, 제1 트리는 대상 피처에 기초한 2개의 노드 분기들을 가지고, 제2 트리는 대상 피처에 기초한 1개의 노드 분기를 가지며, 제3 트리는 대상 피처에 기초한 3개의 노드 분기들을 가지는 경우, 6으로 계산될 수 있다.

[0092] 일 실시예에 따르면, 프로세서는, 게인에 기초하여 임시 부분 입력 데이터의 피처들 중 일부를 선택할 수 있다. 프로세서는, 트레이닝된 트리 기반의 앙상블 모델에 대하여 임시 부분 입력 데이터의 각 피처의 게인을 계산할 수 있다. 프로세서는, 게인에 기초하여 임시 부분 입력 데이터의 피처 중 적어도 하나의 피처를 선택할 수 있다. 임시 부분 입력 데이터의 피처 중에서, 선택된 피처는, 나머지 피처보다 큰 게인을 가질 수 있다.

[0093] 일 실시예에 따르면, 프로세서는, 게인 및 커버에 기초하여 임시 부분 입력 데이터의 피처들 중 일부를 선택할 수 있다. 프로세서는, 트레이닝된 트리 기반의 앙상블 모델에 대하여 임시 부분 입력 데이터의 각 피처의 게인을 계산할 수 있다. 프로세서는, 게인에 기초하여 임시 부분 입력 데이터의 피처 중 적어도 하나의 피처를 선택함으로써 게인에 기초한 피처 세트(feature set)를 획득할 수 있다. 임시 부분 입력 데이터의 피처 중에서, 게인에 기초한 피처 세트의 피처는, 나머지 피처보다 큰 게인을 가질 수 있다. 프로세서는, 트레이닝된 트리 기반의 앙상블 모델에 대하여 임시 부분 입력 데이터의 각 피처의 커버를 계산할 수 있다. 프로세서는, 커버에 기초하여 임시 부분 입력 데이터의 피처 중 적어도 하나의 피처를 선택함으로써 커버에 기초한 피처 세트(feature set)를 획득할 수 있다. 임시 부분 입력 데이터의 피처 중에서, 커버에 기초한 피처 세트의 피처는, 나머지 피처보다 큰 커버를 가질 수 있다. 프로세서는, 임시 부분 입력 데이터의 피처 중 게인에 기초한 피처 세트 및 커버에 기초한 피처 세트의 합집합에 포함된 피처를 선택할 수 있다. 프로세서는, 게인에 기초한 피처 세트 및 커버에 기초한 피처 세트에 모두 포함된 피처를 한 번 선택하도록 중복된 피처에 대하여 처리할 수 있다.

[0094] 단계(530)에서, 프로세서는 선택된 적어도 하나의 피처를 가지는 데이터로 임시 부분 입력 데이터를 업데이트할 수 있다. 프로세서는, 선택된 일부 피처들을 가지는 데이터로 임시 부분 입력 데이터를 변경함으로써, 임시 부분 입력 데이터의 피처의 개수를 감소시킬 수 있다.

[0095] 단계(540)에서, 프로세서는 업데이트된 임시 부분 입력 데이터에 기초하여, 트리 기반의 앙상블 모델을 업데이트할 수 있다. 프로세서는, 업데이트된 임시 부분 입력 데이터에 적용됨으로써 임시 출력 데이터를 산출하는 앙상블 모델로 트리 기반의 앙상블 모델을 업데이트할 수 있다. 트리 기반의 앙상블 모델은, 임시 부분 입력 데이터의 피처의 개수가 감소함에 따라, 업데이트된 임시 부분 입력 데이터에 적용됨으로써 임시 출력 데이터를 산출하도록 트레이닝될 수 있다.

[0096] 일 실시예에 따르면, 프로세서는, 업데이트된 임시 부분 입력 데이터 및 트리 기반의 앙상블 모델에 기초하여, 트리 기반의 앙상블 모델을 트레이닝시키는 동작, 및 피처를 선택하는 동작을 반복할 수 있다. 프로세서는, 임시 부분 입력 데이터의 피처의 개수 및/또는 트리 기반의 앙상블 모델의 성능에 기초하여, 트리 기반의 앙상블 모델을 트레이닝시키는 동작 및 피처를 선택하는 동작을 중단할 수 있다. 프로세서는 임시 부분 입력 데이터의 피처들을 부분 입력 데이터의 피처로 결정할 수 있다. 예를 들어, 프로세서는, 임시 부분 입력 데이터의 피처의 개수가 미리 결정된 개수 이하인 경우, 임시 부분 입력 데이터의 피처들을 부분 입력 데이터의 피처로 선택할 수 있다. 예를 들어, 프로세서는, 기존 트리 기반의 앙상블 모델의 성능이 업데이트된 트리 기반의 앙상블 모델의 성능보다 저하되는 경우, 기존 트리 기반의 앙상블 모델이 적용되는 임시 부분 입력 데이터의 피처들을 부분 입력 데이터의 피처로 선택할 수 있다.

[0098] 도 6은 다양한 실시예들 및 비교 실시예에 따른 시퀀스 모델의 성능을 비교하기 위한 도면이다.

[0099] 도 6에서, 비교 실시예 및 일 실시예들에 따른 시퀀스 모델들은, 입력 데이터에 적용됨으로써 출력 데이터를 산출하는 모델일 수 있다. 입력 데이터는 2개의 부분 입력 데이터를 포함하고, 각 부분 입력 데이터는 12시간의 시간 길이를 가지는 시간대에 대응할 수 있다. 출력 데이터는, 입력 시간대의 종료 시각으로부터 12시간 이후의 시각까지의 시간대에 대응할 수 있다. 비교 실시예 및 일 실시예들에 따른 시퀀스 모델들은, 부분 입력 데이터의 피처에서 주요한 차이점을 가질 수 있다.

[0100] 비교 실시예에 따른 시퀀스 모델의 부분 입력 데이터는 의료 데이터의 복수의 피처들을 가질 수 있다. 다시 말해, 비교 실시예에 따른 시퀀스 모델의 부분 입력 데이터는, 의료 데이터의 복수의 피처들 중 선택된 피처들을 가지는 부분 입력 데이터 대신에, 의료 데이터의 복수의 피처들 모두를 가질 수 있다.

[0101] 일 실시예 1에 따른 시퀀스 모델의 부분 입력 데이터는 의료 데이터의 복수의 피처들 중 선택된 피처를 가질 수 있다. 일 실시예 1에 따른 시퀀스 모델의 부분 입력 데이터의 피처는, 임시 부분 입력 데이터 및 트리 기반의 앙상블 모델의 업데이트를 반복함으로써 선택될 수 있다. 일 실시예 1에 따른 시퀀스 모델의 부분 입력 데이터의 피처는, 시퀀스 모델의 성능에 기초하여 임시 부분 입력 데이터 및 트리 기반의 앙상블 모델의 업데이트를 중단함으로써, 93개의 피처들로 결정될 수 있다.

[0102] 일 실시예 2에 따른 시퀀스 모델의 부분 입력 데이터는 의료 데이터의 복수의 피처들 중 게인(gain)에 기초하여 선택된 50개의 피처들을 가질 수 있다.

[0103] 일 실시예 3에 따른 시퀀스 모델의 부분 입력 데이터는 의료 데이터의 복수의 피처들 중 게인(gain)에 기초하여 선택된 10개의 피처들을 가질 수 있다.

[0104] 일 실시예 4에 따른 시퀀스 모델의 부분 입력 데이터의 피처는, 게인 및 커버에 기초하여 선택될 수 있다. 구체적으로, 일 실시예 4에 따른 시퀀스 모델의 부분 입력 데이터는, 게인에 기초하여 선택된 10개의 피처들 및 커버에 기초하여 선택된 10개의 피처들을 가질 수 있다. 결과적으로, 일 실시예 4에 따른 시퀀스 모델의 부분 입력 데이터는, 게인 및 커버에 기초하여 선택된 20개의 피처들을 가질 수 있다.

[0105] 도 6에서 나타난 바와 같이, 일 실시예들에 따른 시퀀스 모델들 각각은, 비교 실시예에 따른 시퀀스 모델보다 평균 제곱 오차(Mean Squared Error; MSE), 평균 절대 오차(Mean Absolute Error; MAE), 결정 계수(R-Squared), 및 정확도(accuracy)의 측면에서 모두 우수한 성능을 가질 수 있다.

[0106] 또한, 일 실시예들에 따른 시퀀스 모델들 중에서도, 게인 및 커버에 기초하여 피처를 선택하는 일 실시예 4에 따른 시퀀스 모델이, 게인에 기초하여 피처를 일 실시예 1 내지 3에 따른 시퀀스 모델들보다, 평균 제곱 오차(Mean Squared Error; MSE), 평균 절대 오차(Mean Absolute Error; MAE), 결정 계수(R-Squared), 및 정확도(accuracy)의 측면에서 모두 우수한 성능을 가질 수 있다.

[0108] 도 7은 다양한 실시예들 및 비교 실시예에 따른 시퀀스 모델의 성능을 비교하기 위한 도면이다.

[0109] 도 6에서, 비교 실시예 및 일 실시예들에 따른 시퀀스 모델들은, 입력 데이터에 적용됨으로써 출력 데이터를 산출하는 모델일 수 있다. 입력 데이터는 2개의 부분 입력 데이터를 포함하고, 각 부분 입력 데이터는 12시간의 시간 길이를 가지는 시간대에 대응할 수 있다. 출력 데이터는, 출력 시간대에 대응할 수 있고, 출력 시간대의 시작 시각은, 입력 시간대의 종료 시각으로부터 12시간 이후의 시각이고, 출력 시간대의 종료 시각은, 출력 시간대의 시작 시각으로부터 12시간 이후의 시각일 수 있다. 비교 실시예 및 일 실시예들에 따른 시퀀스 모델들은, 부분 입력 데이터의 피처에서 주요한 차이점을 가질 수 있다.

[0110] 비교 실시예에 따른 시퀀스 모델의 부분 입력 데이터는 의료 데이터의 복수의 피처들을 가질 수 있다. 다시 말해, 비교 실시예에 따른 시퀀스 모델의 부분 입력 데이터는, 의료 데이터의 복수의 피처들 중 선택된 피처들을 가지는 부분 입력 데이터 대신에, 의료 데이터의 복수의 피처들 모두를 가질 수 있다.

[0111] 일 실시예 1에 따른 시퀀스 모델의 부분 입력 데이터는 의료 데이터의 복수의 피처들 중 선택된 피처를 가질 수 있다. 일 실시예 1에 따른 시퀀스 모델의 부분 입력 데이터의 피처는, 임시 부분 입력 데이터 및 트리 기반의 앙상블 모델의 업데이트를 반복함으로써 선택될 수 있다. 일 실시예 1에 따른 시퀀스 모델의 부분 입력 데이터의 피처는, 시퀀스 모델의 성능에 기초하여 임시 부분 입력 데이터 및 트리 기반의 앙상블 모델의 업데이트를 중단함으로써, 93개의 피처들로 결정될 수 있다.

[0112] 일 실시예 2에 따른 시퀀스 모델의 부분 입력 데이터는 의료 데이터의 복수의 피처들 중 게인(gain)에 기초하여 선택된 50개의 피처들을 가질 수 있다.

[0113] 일 실시예 3에 따른 시퀀스 모델의 부분 입력 데이터는 의료 데이터의 복수의 피처들 중 게인(gain)에 기초하여 선택된 10개의 피처들을 가질 수 있다.

[0114] 도 7에서 나타난 바와 같이, 일 실시예들에 따른 시퀀스 모델들 각각은, 비교 실시예에 따른 시퀀스 모델보다 평균 제곱 오차(Mean Squared Error; MSE), 평균 절대 오차(Mean Absolute Error; MAE), 결정 계수(R-Squared), 및 정확도(accuracy)의 측면에서 모두 우수한 성능을 가질 수 있다.

[0115] 또한, 일 실시예들에 따른 시퀀스 모델들 중에서도, 게인에 기초하여 10개의 피처들을 선택하는 일 실시예 3에 따른 시퀀스 모델이, 게인에 기초하여 더 많은 피처들을 선택하는 일 실시예 1 및 2에 따른 시퀀스 모델들보다, 평균 제곱 오차(Mean Squared Error; MSE), 평균 절대 오차(Mean Absolute Error; MAE), 결정 계수(R-Squared), 및 정확도(accuracy)의 측면에서 모두 우수한 성능을 가질 수 있다.

[0117] 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 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.

[0118] 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.

[0119] 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.

[0120] 위에서 설명한 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 또는 복수의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

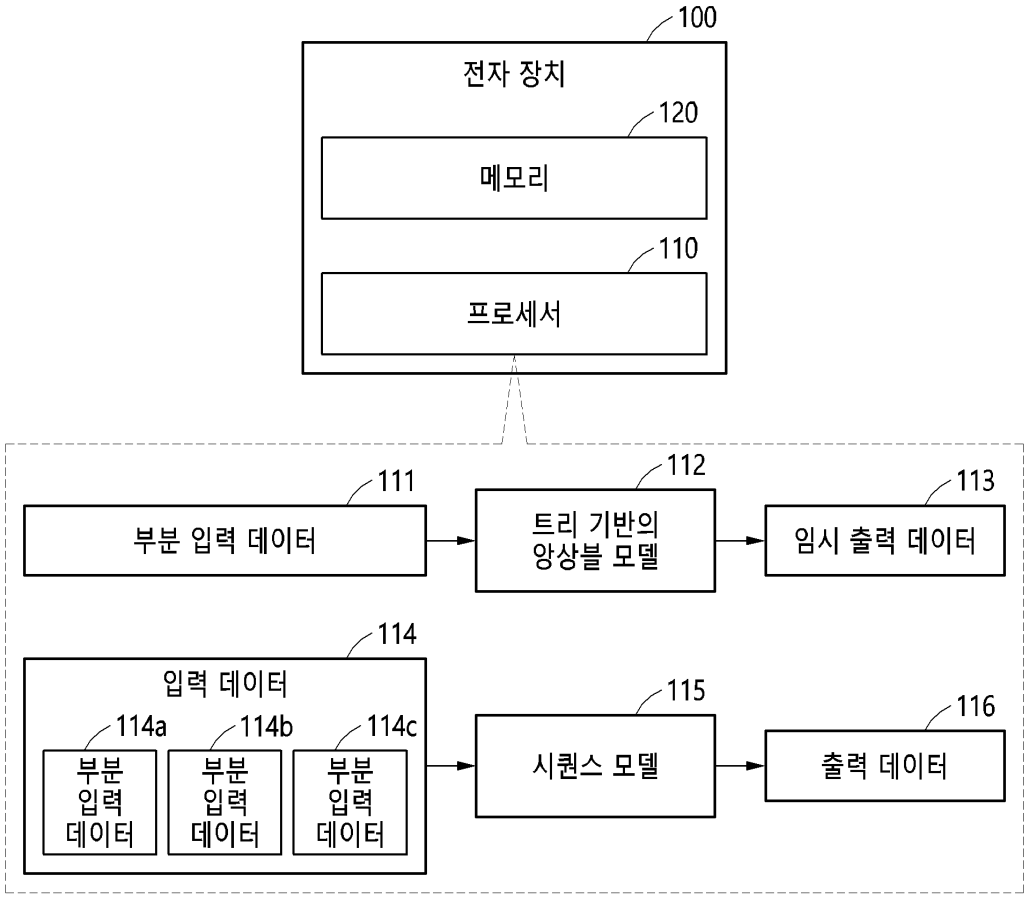
[0121] 본 문서에서, "A 또는 B", "A 및 B 중 적어도 하나", "A 또는 B 중 적어도 하나", "A, B 또는 C", "A, B 및 C 중 적어도 하나", 및 "A, B, 또는 C 중 적어도 하나"와 같은 문구들 각각은 그 문구들 중 해당하는 문구에 함께 나열된 항목들 중 어느 하나, 또는 그들의 모든 가능한 조합을 포함할 수 있다.

[0122] 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.

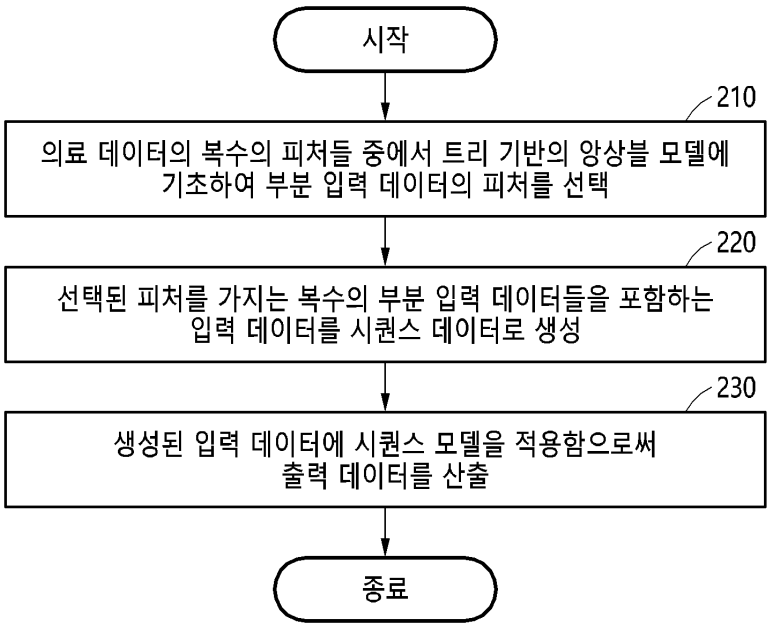
[0123] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

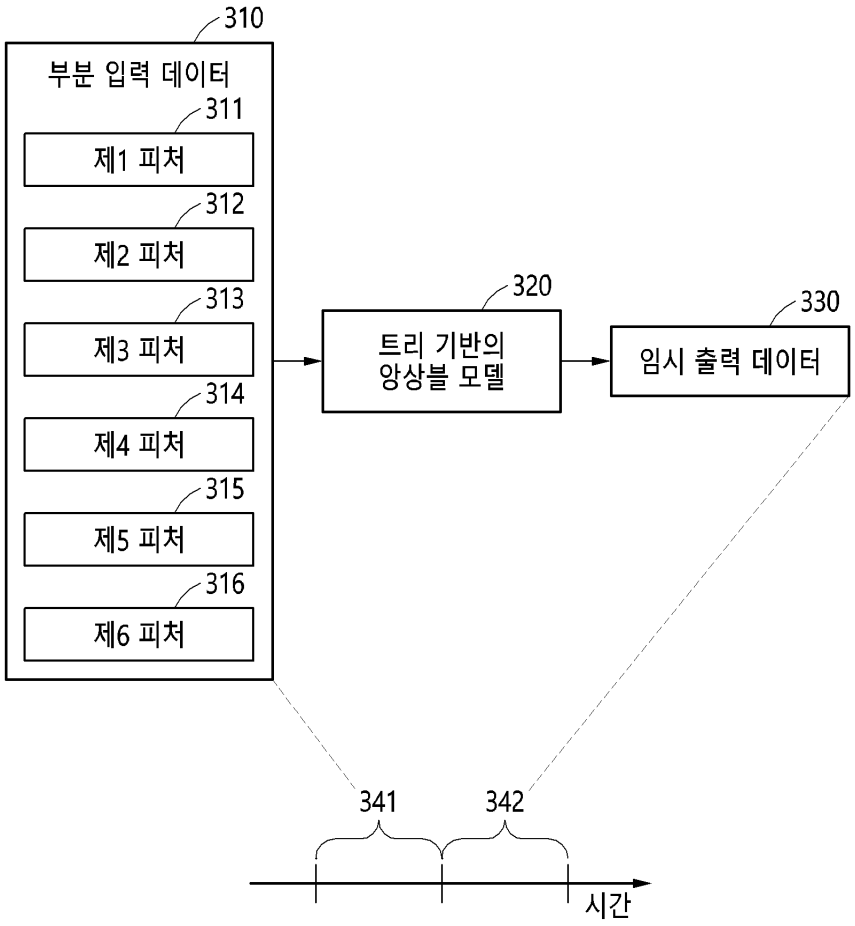
도면1


100
전자 장치
120
메모리
110
프로세서
111
부분 입력 데이터
112
트리 기반의
앙상블 모델
113
임시 출력 데이터
114
입력 데이터
114a
부분
입력
데이터
114b
부분
입력
데이터
114c
부분
입력
데이터
115
시퀀스 모델
116
출력 데이터

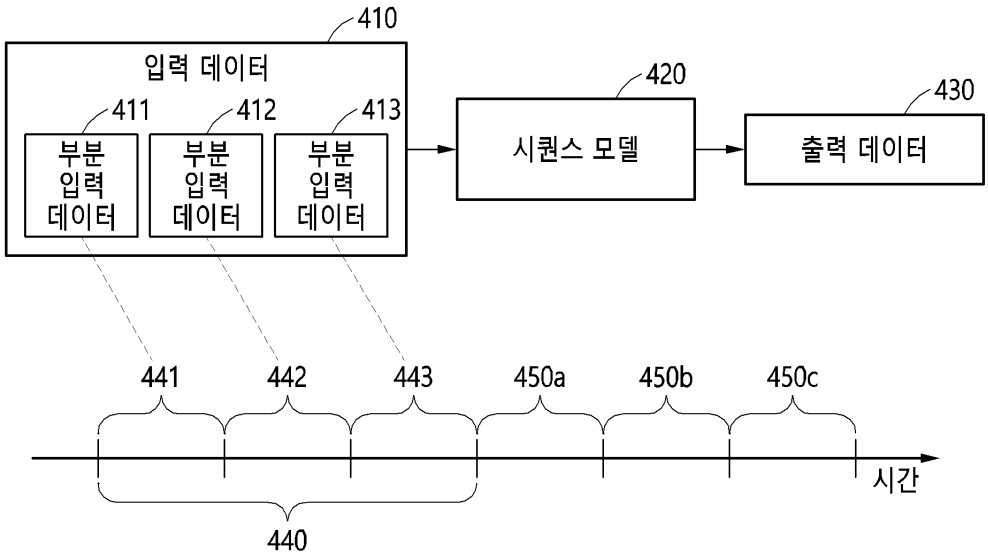
도면2


시작
210
의료 데이터의 복수의 피처들 중에서 트리 기반의 앙상블 모델에
기초하여 부분 입력 데이터의 피처를 선택
220
선택된 피처를 가지는 복수의 부분 입력 데이터들을 포함하는
입력 데이터를 시퀀스 데이터로 생성
230
생성된 입력 데이터에 시퀀스 모델을 적용함으로써
출력 데이터를 산출
종료

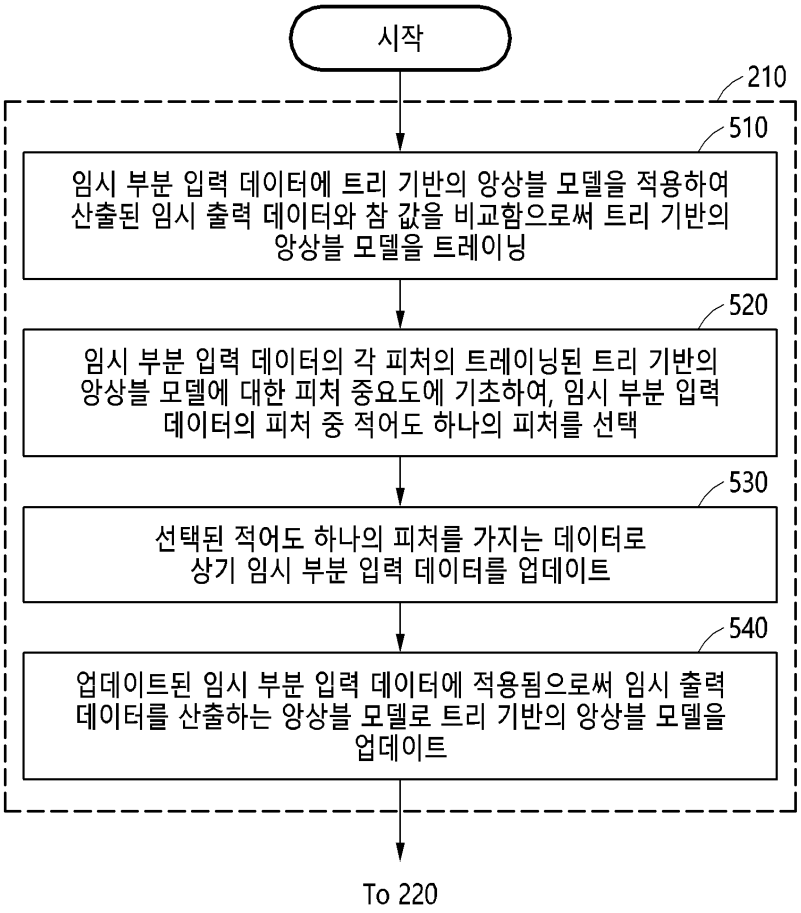
도면3


310
부분 입력 데이터
311
제1 피처
312
제2 피처
313
제3 피처
314
제4 피처
315
제5 피처
316
제6 피처
320
트리 기반의
앙상블 모델
330
임시 출력 데이터
341
342
시간

도면4


410
입력 데이터
411
412
413
부분
입력
데이터
부분
입력
데이터
부분
입력
데이터
420
시퀀스 모델
430
출력 데이터
441
442
443
450a
450b
450c
시간
440

도면5



도면6

600

Number of Features	MSE	MAE	R-Squared	Accuracy
비교 실시예	77.48	5.61	0.92	0.82
일 실시예 1	47.23	4.43	0.95	0.86
일 실시예 2	43.04	4.29	0.96	0.87
일 실시예 3	30.28	3.43	0.97	0.89
일 실시예 4	24.809	3.014	0.976	0.903

도면7

700

Number of Features	MSE	MAE	R-Squared	Accuracy
비교 실시예	126.89	7.73	0.88	0.75
일 실시예 1	98.909	6.757	0.904	0.791
일 실시예 2	83.675	6.125	0.919	0.799
일 실시예 3	53.583	4.591	0.948	0.852