(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0018281
(43) 공개일자 2025년02월05일

(51) 국제특허분류(Int. Cl.)
G16H 50/30 (2018.01) *G06N 3/04* (2023.01)
G06N 3/08 (2023.01) *G16H 10/60* (2018.01)
G16H 50/50 (2018.01) *G16H 50/70* (2018.01)
(52) CPC특허분류
G16H 50/30 (2018.01)
G06N 3/04 (2023.01)
(21) 출원번호 10-2023-0098470
(22) 출원일자 2023년07월27일
심사청구일자 2023년07월27일

(71) 출원인
재단법인 아산사회복지재단
서울특별시 송파구 올림픽로43길 88 (풍납동)
울산대학교 산학협력단
울산광역시 남구 대학로 93(무거동)
(72) 발명자
김성훈
서울시 송파구 풍납동 올림픽로69길
박형준
서울특별시 중구 다산로 32 신당동 남산타운25동 1804호
(74) 대리인
특허법인비엘티

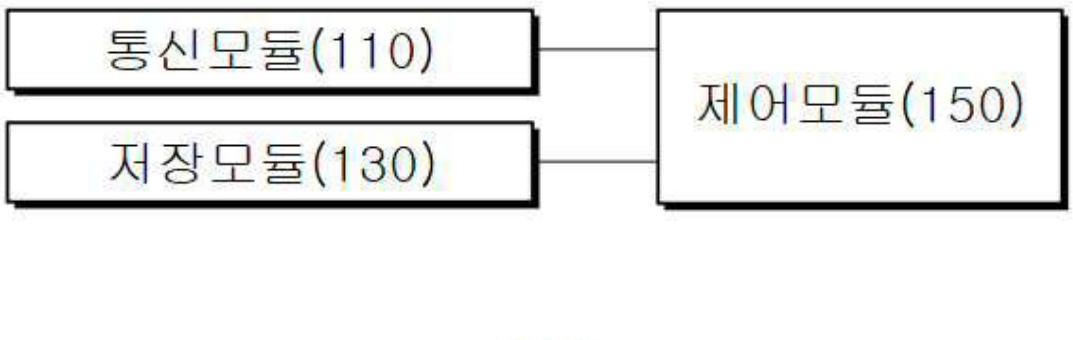
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치 및 방법**

(57) 요 약

본 개시는 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치 및 방법에 관한 것으로, 외부 장치와 통신을 수행하는 통신모듈; 적어도 하나의 환자에 대한 위험도 및 악화 예측을 위한 적어도 하나의 프로세스를 저장하는 저장모듈; 및 상기 적어도 하나의 프로세스를 기반으로 상기 적어도 하나의 환자에 대한 위험도 및 악화 예측을 위한 동작을 수행하는 제어모듈을 포함하며, 상기 제어모듈은, 상기 저장모듈 또는 별도의 데이터베이스에서 적어도 하나의 환자 중 특정 환자에 대한 EMR 데이터를 획득하고, 제로 패딩(zero padding)을 적용하여 상기 획득된 EMR 데이터에 대한 전처리를 수행한 후, 상기 전처리된 EMR 데이터를 기반으로 기 설정된 인공지능 기반의 학습모델을 이용하여 상기 특정 환자에 대한 위험도 및 악화 예측에 대한 결과정보를 생성할 수 있다.

대 표 도 - 도1



(52) CPC특허분류
G06N 3/08 (2023.01)
G16H 10/60 (2021.08)
G16H 50/50 (2018.01)
G16H 50/70 (2018.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1465038587
과제번호	HR20C0026010023
부처명	보건복지부
과제관리(전문)기관명	한국보건산업진흥원
연구사업명	연구중심병원 육성 R&D
연구과제명	고위험 환자 최적진료를 위한 차세대 융합 케어기술개발 및 사업화
기 여 율	1/1
과제수행기관명	서울아산병원
연구기간	2023.01.01 ~ 2023.12.31

명 세 서

청구범위

청구항 1

외부 장치와 통신을 수행하는 통신모듈;

적어도 하나의 환자에 대한 위험도 및 악화 예측을 위한 적어도 하나의 프로세스를 저장하는 저장모듈; 및

상기 적어도 하나의 프로세스를 기반으로 상기 적어도 하나의 환자에 대한 위험도 및 악화 예측을 위한 동작을 수행하는 제어모듈을 포함하며,

상기 제어모듈은,

상기 저장모듈 또는 별도의 데이터베이스에서 적어도 하나의 환자 중 특정 환자에 대한 EMR 데이터를 획득하고, 제로 패딩(zero padding)을 적용하여 상기 획득된 EMR 데이터에 대한 전처리를 수행한 후, 상기 전처리된 EMR 데이터를 기반으로 기 설정된 인공지능 기반의 학습모델을 이용하여 상기 특정 환자에 대한 위험도 및 악화 예측에 대한 결과정보를 생성하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치.

청구항 2

제1항에 있어서,

상기 제어모듈은,

기 설정된 기준정보를 기반으로 최적의 데이터 셋 사이즈를 결정하고, 상기 이상치가 제거된 EMR 데이터를 이용하여 상기 결정된 데이터 셋 사이즈에 따라 특정 항목에 대한 데이터 셋을 구축하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치.

청구항 3

제2항에 있어서,

상기 EMR 데이터는,

의료진에 의하여 입력되는 데이터로서, 진료기록, 검사, 투약, 증상 및 처치 처방 중 적어도 하나에 대한 데이터를 포함하고,

상기 검사에 대한 데이터는 활력징후(vital sign)에 대한 측정 데이터 및 혈액검사에 따른 데이터 중 적어도 하나의 수치형 데이터를 포함하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치.

청구항 4

제1항에 있어서,

상기 제어모듈은,

상기 획득된 EMR 데이터에 대한 전처리를 수행 시에,

상기 획득된 EMR 데이터에 포함된 적어도 하나의 항목 중 특정 항목에 대한 이상치(Outlier)를 확인하고, 상기 확인된 이상치를 제거한 후, 기 설정된 기준정보를 기반으로 최적의 데이터 셋 사이즈를 결정하고, 상기 이상치가 제거된 EMR 데이터를 이용하여 상기 결정된 데이터 셋 사이즈에 따라 특정 항목에 대한 데이터 셋을 구축하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치.

청구항 5

제4항에 있어서,

상기 기 설정된 기준정보는,

질병, 증상 및 환자 상태에 따라 최적의 데이터 셋 사이즈가 맵핑된 저장된 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치.

청구항 6

제5항에 있어서,

상기 제어모듈은,

상기 획득된 EMR 데이터에 대한 전처리를 수행 시에,

상기 결정된 데이터 셋 사이즈에 따라 부족한 값은 제로값을 추가함으로써 상기 제로 패딩을 적용하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치.

청구항 7

제6항에 있어서,

상기 제어모듈은,

상기 획득된 EMR 데이터에 대한 전처리를 수행 시에,

동일한 환자에 대한 상기 특정 항목의 값을 그룹화하여 묶고, 중앙값(median) 및 차이값(difference)로 확인하고, 상기 확인된 중앙값 및 차이값을 각각에 대해 기 설정된 임계치와 비교하고, 상기 기 설정된 임계치를 벗어나는 값을 제거하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치.

청구항 8

제1항에 있어서,

상기 제어모듈은,

상기 전처리를 수행하기 이전에, 상기 획득된 EMR 데이터에 대해 결측치가 발생하는지 여부를 판단하여 상기 위험도 및 악화 예측에 적합한 환자 데이터인지 여부를 판단하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치.

청구항 9

장치에 의해 수행되는, 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법에 있어서,

저장모듈 또는 별도의 데이터베이스에서 적어도 하나의 환자 중 특정 환자에 대한 EMR 데이터를 획득하는 단계;

제로 패딩(zero padding)을 적용하여 상기 획득된 EMR 데이터에 대한 전처리를 수행하는 단계; 및

상기 전처리된 EMR 데이터를 기반으로 기 설정된 인공지능 기반의 학습모델을 이용하여 상기 특정 환자에 대한 위험도 및 악화 예측에 대한 결과정보를 생성하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법.

청구항 10

제9항에 있어서,

상기 획득된 EMR 데이터에 대한 전처리를 수행하는 단계는,

기 설정된 기준정보를 기반으로 최적의 데이터 셋 사이즈를 결정하고, 상기 이상치가 제거된 EMR 데이터를 이용하여 상기 결정된 데이터 셋 사이즈에 따라 특정 항목에 대한 데이터 셋을 구축하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법.

청구항 11

제10항에 있어서,

상기 EMR 데이터는,

의료진에 의하여 입력되는 데이터로서, 진료기록, 검사, 투약, 증상 및 처치 처방 중 적어도 하나에 대한 데이터를 포함하고,

상기 검사에 대한 데이터는 활력징후(vital sign)에 대한 측정 데이터 및 혈액검사에 따른 데이터 중 적어도 하나의 수치형 데이터를 포함하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법.

청구항 12

제9항에 있어서,

상기 획득된 EMR 데이터에 대한 전처리를 수행하는 단계는,

상기 획득된 EMR 데이터에 포함된 적어도 하나의 항목 중 특정 항목에 대한 이상치(Outlier)를 확인하고, 상기 확인된 이상치를 제거한 후, 기 설정된 기준정보를 기반으로 최적의 데이터 셋 사이즈를 결정하고, 상기 이상치가 제거된 EMR 데이터를 이용하여 상기 결정된 데이터 셋 사이즈에 따라 특정 항목에 대한 데이터 셋을 구축하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법.

청구항 13

제12항에 있어서,

상기 기 설정된 기준정보는,

질병, 증상 및 환자 상태에 따라 최적의 데이터 셋 사이즈가 맵핑된 저장된 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법.

청구항 14

제13항에 있어서,

상기 획득된 EMR 데이터에 대한 전처리를 수행하는 단계는,

상기 결정된 데이터 셋 사이즈에 따라 부족한 값은 제로값을 추가함으로써 상기 제로 패딩을 적용하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법.

청구항 15

제14항에 있어서,

상기 획득된 EMR 데이터에 대한 전처리를 수행하는 단계는,

동일한 환자에 대한 상기 특정 항목의 값을 그룹화하여 묶고, 중앙값(median) 및 차이값(difference)로 확인하고, 상기 확인된 중앙값 및 차이값을 각각에 대해 기 설정된 임계치와 비교하고, 상기 기 설정된 임계치를 벗어나는 값을 제거하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법.

청구항 16

제9항에 있어서,

상기 제어모듈은,

상기 전처리를 수행하는 단계 이전에,

상기 획득된 EMR 데이터에 대해 결측치가 발생하는지 여부를 판단하여 상기 위험도 및 악화 예측에 적합한 환자 데이터인지 여부를 판단하는 단계를 더 포함하는 것을 특징으로 하는,

환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법.

청구항 17

컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램으로서, 상기 컴퓨터 프로그램은 하나 이상의 프로세서에서 실행되는 경우, 데이터 처리 장치에 의해 수행되는 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법을 수행하기 위한 이하의 동작들을 수행하도록 하며, 상기 동작들은:

저장모듈 또는 별도의 데이터베이스에서 적어도 하나의 환자 중 특정 환자에 대한 EMR 데이터를 획득하는 동작;

제로 패딩(zero padding)을 적용하여 상기 획득된 EMR 데이터에 대한 전처리를 수행하는 동작; 및

상기 전처리된 EMR 데이터를 기반으로 기 설정된 인공지능 기반의 학습모델을 이용하여 상기 특정 환자에 대한 위험도 및 악화 예측에 대한 결과정보를 생성하는 동작;을 포함하는 것을 특징으로 하는,

컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램.

청구항 18

제17항에 있어서,

상기 획득된 EMR 데이터에 대한 전처리를 수행하는 동작은,

기 설정된 기준정보를 기반으로 최적의 데이터 셋 사이즈를 결정하고, 상기 이상치가 제거된 EMR 데이터를 이용하여 상기 결정된 데이터 셋 사이즈에 따라 특정 항목에 대한 데이터 셋을 구축하는 것을 특징으로 하는,

컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램.

청구항 19

제18항에 있어서,

상기 EMR 데이터는,

의료진에 의하여 입력되는 데이터로서, 진료기록, 검사, 투약, 증상 및 처치 처방 중 적어도 하나에 대한 데이터를 포함하고,

상기 검사에 대한 데이터는 활력징후(vital sign)에 대한 측정 데이터 및 혈액검사에 따른 데이터 중 적어도 하나의 수치형 데이터를 포함하는 것을 특징으로 하는,

컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램.

청구항 20

제17항에 있어서,

상기 획득된 EMR 데이터에 대한 전처리를 수행하는 동작은,

상기 획득된 EMR 데이터에 포함된 적어도 하나의 항목 중 특정 항목에 대한 이상치(Outlier)를 확인하고, 상기 확인된 이상치를 제거한 후, 기 설정된 기준정보를 기반으로 최적의 데이터 셋 사이즈를 결정하고, 상기 이상치가 제거된 EMR 데이터를 이용하여 상기 결정된 데이터 셋 사이즈에 따라 특정 항목에 대한 데이터 셋을 구축하는 것을 특징으로 하는,

컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기 술 분 야

[0001] 본 개시는 데이터 처리 장치 및 방법에 대한 것으로, 보다 상세하게는, 본 개시는 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 병원정보시스템(Hospital Infromation System, HIS)은 업무효율화와 정보연계를 도모하는 시스템으로서 의료시설(병원 및 일반진료소)에 도입되고 있다. 특히 가장 직접적이고 정확한 의료 데이터가 생성되는 병원은 일찍이 전자의무기록시스템(Electronic Medical Record, EMR) 등을 활용한 이 병원정보시스템을 구축하며 흐름에 발을 맞췄다.

[0003] 환자의 의료기록을 디지털 형식으로 기록하여 보관하는 EMR은 중요한 헬스케어 데이터로 꼽힌다. 이 EMR은 시계열 데이터로, 질병의 악화 및 회복 등에 따라 변화하는 환자 건강상태에 대한 기록이다. 한 시점의 환자 건강상태는 과거 상태의 영향을 받으므로 해당 시점 환자 상태에 영향을 준 최근 건강상태 영역이 존재한다. 미래 시점의 건강 상태는 시간의 흐름에 따른 EMR을 통해 추이를 분석함으로써 예측될 수 있다.

[0004] 기존에는 EMR을 이용하여 환자의 중증도를 예측해 왔으나, 입원환자의 초기 데이터가 없어 일정 기간이 지나야 정확한 중증도를 예측할 수 있었다.

[0005] 따라서, 초기 입원환자의 의료 데이터셋이 부족한 상황에서도 초기 시계열 정보만으로도 환자에 대한 위험도 및 악화 예측을 보다 정확하게 예측할 수 있도록 하는 기술이 개발될 필요가 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국공개특허공보 제10-2022-0115038호 (공개일: 2022년 08월 17일)

발명의 내용

해결하려는 과제

[0007] 본 개시는 초기 입원환자의 부족한 의료 데이터셋을 일정 부분 제로값(0)으로 대체함으로써, 초기 시계열 정보만으로도 중증도 예측 정확도를 향상시킬 수 있도록 하는 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치 및 방법을 제공함에 있다.

[0008] 본 개시가 해결하고자 하는 과제들은 이상에서 언급된 과제로 제한되지 않으며, 언급되지 않은 또 다른 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0009] 상술한 기술적 과제를 달성하기 위한 본 개시의 일 측면에 따른 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치는, 외부 장치와 통신을 수행하는 통신모듈; 적어도 하나의 환자에 대한 위험도 및 악화 예측을 위한 적어도 하나의 프로세스를 저장하는 저장모듈; 및 상기 적어도 하나의 프로세스를 기반으로 상기 적어도 하나의 환자에 대한 위험도 및 악화 예측을 위한 동작을 수행하는 제어모듈을 포함하며, 상기 제어모듈은, 상기 저장모듈 또는 별도의 데이터베이스에서 적어도 하나의 환자 중 특정 환자에 대한 EMR 데이터를 획득하고, 제로 패딩(zero padding)을 적용하여 상기 획득된 EMR 데이터에 대한 전처리를 수행한 후, 상기 전처리된 EMR 데이터를 기반으로 기 설정된 인공지능 기반의 학습모델을 이용하여 상기 특정 환자에 대한 위험도 및 악화 예측에 대한 결과정보를 생성할 수 있다.

[0010] 한편, 본 개시의 일 측면에 따른 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법은, 저장모듈 또는 별도의 데이터베이스에서 적어도 하나의 환자 중 특정 환자에 대한 EMR 데이터를 획득하는 단계; 제로 패딩(zero padding)을 적용하여 상기 획득된 EMR 데이터에 대한 전처리를 수행하는 단계; 및 상기 전처리된 EMR 데

이터를 기반으로 기 설정된 인공지능 기반의 학습모델을 이용하여 상기 특정 환자에 대한 위험도 및 악화 예측에 대한 결과정보를 생성할 수 있다.

[0011] 이 외에도, 본 개시를 구현하기 위한 방법을 실행하기 위한 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램이 더 제공될 수 있다.

[0012] 이 외에도, 본 개시를 구현하기 위한 방법을 실행하기 위한 컴퓨터 프로그램을 기록하는 컴퓨터 판독 가능한 기록 매체가 더 제공될 수 있다.

발명의 효과

[0013] 본 개시의 전술한 과제 해결 수단에 의하면, 초기 입원환자의 부족한 의료 데이터셋을 일정 부분 제로값(0)으로 대체함으로써, 초기 시계열 정보만으로도 중증도 예측 정확도를 향상시킬 수 있도록 한다.

[0014] 본 개시의 효과들은 이상에서 언급된 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0015] 도 1은 본 개시의 일 실시예에 따른 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치의 구성을 나타내는 블록도

도 2는 본 개시의 일 실시예에 따른 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법을 나타내는 순서도

도 3은 본 개시의 일 실시예에 따라 EMR 데이터에 대한 전처리를 수행하는 동작을 구체적으로 나타내는 순서도

도 4는 본 개시의 일 실시예에 따른 전처리 동작을 통해 EMR 데이터에 포함된 적어도 하나의 항목 중 키 항목에 대한 데이터 셋을 구축하는 일 예시를 나타내는 도면

도 5는 본 개시의 일 실시예에 따른 전처리 동작을 통해 EMR 데이터에 포함된 적어도 하나의 항목 중 체중 항목에 대한 데이터 셋을 구축하는 일 예시를 나타내는 도면

도 6은 본 개시의 일 실시예에 따른 기 설정된 인공지능 기반의 학습모델을 이용하여 서로 다른 데이터 셋 사이즈를 갖는 데이터 셋을 기반으로 예측된 결과의 일 예시를 나타내는 도면

도 7은 본 개시의 일 실시예에 따른 인공지능 기반의 학습모델의 기본 구조를 나타내는 도면

도 8은 본 개시에 따른 인공지능 기반의 학습모델에 의해 생성된 환자에 대한 위험도 및 악화 예측에 대한 결과정보의 일 예시를 나타내는 도면

발명을 실시하기 위한 구체적인 내용

[0016] 본 개시의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 개시는 이하에서 개시되는 실시예들에 제한되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 개시가 완전하도록 하고, 본 개시가 속하는 기술 분야의 통상의 기술자에게 본 개시의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 개시는 청구항의 범주에 의해 정의될 뿐이다.

[0017] 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 개시를 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 "포함한다(comprises)" 및/또는 "포함하는(comprising)"은 언급된 구성요소 외에 하나 이상의 다른 구성요소의 존재 또는 추가를 배제하지 않는다. 명세서 전체에 걸쳐 동일한 도면 부호는 동일한 구성 요소를 지칭하며, "및/또는"은 언급된 구성요소들의 각각 및 하나 이상의 모든 조합을 포함한다. 비록 "제1", "제2" 등이 다양한 구성요소들을 서술하기 위해서 사용되나, 이들 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소를 다른 구성요소와 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제1 구성요소는 본 개시의 기술적 사상 내에서 제2 구성요소일 수도 있음은 물론이다.

[0018] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 개시가 속하는 기술 분야의 통상의 기술자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또한, 일반적으로 사용되

는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.

[0019] 본 개시 전체에 걸쳐 동일 참조 부호는 동일 구성요소를 지칭한다. 본 개시가 실시예들의 모든 요소들을 설명하는 것은 아니며, 본 개시가 속하는 기술분야에서 일반적인 내용 또는 실시예들 간에 중복되는 내용은 생략한다. 명세서에서 사용되는 "부" 또는 “모듈”이라는 용어는 소프트웨어, FPGA 또는 ASIC과 같은 하드웨어 구성요소를 의미하며, "부" 또는 “모듈”은 어떤 역할들을 수행한다. 그렇지만 "부" 또는 “모듈”은 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. "부" 또는 “모듈”은 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 "부" 또는 “모듈”은 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다. 구성요소들과 "부" 또는 “모듈”들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 "부" 또는 “모듈”들로 결합되거나 추가적인 구성요소들과 "부" 또는 “모듈”들로 더 분리될 수 있다.

[0020] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 직접적으로 연결되어 있는 경우뿐 아니라, 간접적으로 연결되어 있는 경우를 포함하고, 간접적인 연결은 무선 통신망을 통해 연결되는 것을 포함한다.

[0021] 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0022] 명세서 전체에서, 어떤 부재가 다른 부재 "상에" 위치하고 있다고 할 때, 이는 어떤 부재가 다른 부재에 접해 있는 경우뿐 아니라 두 부재 사이에 또 다른 부재가 존재하는 경우도 포함한다.

[0023] 제1, 제2 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하기 위해 사용되는 것으로, 구성요소가 전술된 용어들에 의해 제한되는 것은 아니다.

[0024] 단수의 표현은 문맥상 명백하게 예외가 있지 않는 한, 복수의 표현을 포함한다.

[0025] 각 단계들에 있어 식별부호는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단계들의 순서를 설명하는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않는 이상 명기된 순서와 다르게 실시될 수 있다.

[0026] 이하 설명에서 사용되는 용어에 대하여 정의하면 하기와 같다.

[0027] 본 명세서에서 'EMR(Electronic Medical Record)'는, 병원에 내방한 환자에 대한 진료기록을 기존 종이에 기록하던 것을 컴퓨터를 이용하여 전자적 형태로 저장한 전자의무기록을 의미할 수 있으며, 의료행위를 정확하고 효율적으로 보조하기 위해 운영되고 있다.

[0028] 본 명세서에서 '데이터 처리 장치(100)'로 한정하여 설명하였으나, 이는 EMR 데이터를 이용하여 중환자의 위험도 및 악화 예측을 수행하고자 하는 사용자의 단말(의료진 단말)로서, 연산처리를 수행할 수 있는 다양한 장치들을 모두 포함할 수 있다. 즉, 데이터 처리 장치(100)는 서버, 컴퓨터, 서버 및/또는 휴대용 단말기를 더 포함하거나, 또는 그 중 어느 하나의 형태가 될 수 있으며, 이를 한정하지 않는다.

[0029] 여기에서, 상기 컴퓨터는 예를 들어, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(desktop), 랩톱(laptop), 태블릿 PC, 슬레이트 PC 등을 포함할 수 있다.

[0030] 상기 서버는 외부 장치와 통신을 수행하여 정보를 처리하는 서버로써, 애플리케이션 서버, 컴퓨팅 서버, 데이터베이스 서버, 파일 서버, 게임 서버, 메일 서버, 프록시 서버 및 웹 서버 등을 포함할 수 있다.

[0031] 상기 휴대용 단말기는 예를 들어, 휴대성과 이동성이 보장되는 무선 통신 장치로서, PCS(Personal Communication System), GSM(Global System for Mobile communications), PDC(Personal Digital Cellular), PHS(Personal Handyphone System), PDA(Personal Digital Assistant), IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(W-Code Division Multiple Access), WiBro(Wireless Broadband Internet) 단말, 스마트 폰(Smart Phone) 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치와 시계, 반지, 팔찌, 발찌, 목걸이, 안경, 콘택트 렌즈, 또는 머리 착용형 장치(head-mounted-device(HMD) 등과 같은 웨어러블 장치를 포함할 수 있다.

[0032] 이하 첨부된 도면들을 참고하여 본 개시의 작용 원리 및 실시예들에 대해 설명한다.

[0033] 도 1은 본 개시의 일 실시예에 따른 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치의 구성을 나타내는 블록도이다.

[0034] 도 1을 참조하면, 본 개시의 일 실시예에 따른 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 장치(이하, '데이터 처리 장치'라 칭함)(100)는 통신모듈(110), 저장모듈(130) 및 제어모듈(150)을 포함하여 구성될 수 있다.

[0035] 통신모듈(110)은 적어도 하나의 장치/단말과 적어도 하나의 정보 또는 데이터를 송수신한다. 여기서, 적어도 하나의 장치/단말은 적어도 하나의 환자에 대한 위험도 및 악화 예측에 대한 결과 정보를 제공받고자 하는 장치/단말 이거나, 그 결과 정보를 생성할 수 있도록 하는 각종 데이터/정보를 제공하는 장치/단말일 수 있으며, 그 종류 및 형태를 한정하지 않는다.

[0036] 또한, 이 통신모듈(110)은 그 외 다른 장치들과의 통신을 수행할 수도 있는 것으로, 무선 인터넷 기술들에 따른 통신망에서 무선 신호를 송수신하도록 한다.

[0037] 무선 인터넷 기술로는, 예를 들어 WLAN(Wireless LAN), Wi-Fi(Wireless-Fidelity), Wi-Fi(Wireless Fidelity) Direct, DLNA(Digital Living Network Alliance), WiBro(Wireless Broadband), WiMAX(World Interoperability for Microwave Access), HSDPA(High Speed Downlink Packet Access), HSUPA(High Speed Uplink Packet Access), LTE(Long Term Evolution), LTE-A(Long Term Evolution-Advanced) 등이 있으며, 생성 장치(100)는 앞에서 나열되지 않은 인터넷 기술까지 포함한 범위에서 적어도 하나의 무선 인터넷 기술에 따라 데이터를 송수신하게 된다.

[0038] 근거리 통신(Short range communication)을 위한 것으로서, 블루투스(Bluetooth™), RFID(Radio Frequency Identification), 적외선 통신(Infrared Data Association; IrDA), UWB(Ultra Wideband), ZigBee, NFC(Near Field Communication), Wi-Fi(Wireless-Fidelity), Wi-Fi Direct, Wireless USB(Wireless Universal Serial Bus) 기술 중 적어도 하나를 이용하여, 근거리 통신을 지원할 수 있다. 이러한, 근거리 무선 통신망(Wireless Area Networks)을 생성 장치(100) 및 적어도 하나의 사용자 단말 간 무선 통신을 지원할 수 있다. 이때, 근거리 무선 통신망은 근거리 무선 개인 통신망(Wireless Personal Area Networks)일 수 있다.

[0039] 저장모듈(130)은 적어도 하나의 환자에 대한 위험도 및 악화 예측을 위한 적어도 하나의 프로세스(알고리즘)는 또는 그 프로세스를 재현한 프로그램ㄹ에 대한 데이터를 저장할 수 있다. 뿐만 아니라, 저장모듈(130)은 그 외 다른 동작을 수행하기 위한 프로세스들을 더 저장할 수 있으며, 이를 한정하지 않는다.

[0040] 저장모듈(130)은 적어도 하나의 환자에 대한 인적사항, 기본정보 및/또는 EMR 데이터는 물론, 데이터 처리 장치(100)의 다양한 기능을 지원하는 각종 데이터를 저장할 수 있다. 저장모듈(130)은 데이터 처리 장치(100)에서 구동되는 다수의 응용 프로그램(application program 또는 애플리케이션(application), 생성 장치(100)의 동작을 위한 데이터들, 명령어들을 저장할 수 있다. 이러한 응용 프로그램 중 적어도 일부는, 무선 통신을 통해 외부 서버로부터 다운로드 될 수 있다. 한편, 응용 프로그램은, 저장모듈(130)에 구비된 적어도 하나의 메모리에 저장되고, 데이터 처리 장치(100) 상에 설치되어, 제어모듈(150)을 통해 저장모듈(130)에 저장된 적어도 하나의 프로세서에 의하여 동작(또는 기능)을 수행하도록 구동될 수 있다.

[0041] 한편, 적어도 하나의 메모리는 플래시 메모리 타입(flash memory type), 하드 디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(Random Access Memory, RAM), SRAM(Static Random Access Memory), 롬(Read-Only Memory, ROM), EEPROM(Electrically Erasable Programmable Read-Only Memory), PROM(Programmable Read-Only Memory) 자기 메모리, 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다. 아울러, 메모리는 일시적, 영구적 또는 반영구적으로 정보를 저장할 수 있으며, 내장형 또는 탈착형으로 제공될 수 있다.

[0042] 저장모듈(130)은 적어도 하나의 환자에 대한 위험도 및 악화 예측을 위한 데이터베이스를 구축하거나, 별도의 외부 서버와 연동될 수도 있다.

[0043] 제어모듈(150)은 응용 프로그램과 관련된 동작 외에도, 적어도 하나의 프로세서를 기반으로 데이터 처리 장치(100) 내 모든 구성들을 제어하여 입력 또는 출력되는 신호, 데이터, 정보 등을 처리하거나, 적어도 하나의 메모리에 저장된 명령어, 알고리즘, 응용 프로그램을 실행하여 각종 프로세스를 수행하며, 환자에 대한 위험도 및 악화 예측을 위한 적절한 정보 또는 기능을 제공 또는 처리할 수 있다.

[0044] 구체적으로, 제어모듈(150)은 저장모듈(130) 또는 별도의 데이터베이스에서 적어도 하나의 환자 중 특정 환자에 대한 EMR 데이터를 획득(수집)하고, 제로 패딩(zero padding)을 적용하여 그 획득된 EMR 데이터에 대한 전처리를 수행한 후, 전처리된 EMR 데이터를 기반으로 기 설정된 인공지능 기반의 학습모델을 이용하여 그 특정 환자에 대한 위험도 및 악화 예측에 대한 결과정보를 생성한다.

[0045] 이 제어모듈(150)은 획득된 EMR에 대한 전처리를 수행 시에, 획득된 EMR 데이터에 포함된 적어도 하나의 항목 중 특정 항목에 대한 이상치(Outlier)를 확인하고, 그 확인된 이상치를 제거한다. 이후, 제어모듈(150)은 기 설정된 기준정보를 기반으로 최적의 데이터 셋 사이즈를 결정하고, 앞서 이상치가 제거된 EMR 데이터를 이용하여 그 결정된 데이터 셋 사이즈에 따라 특정 항목에 대한 데이터 셋을 구축한다. 여기서, 데이터 셋을 구축 시에, 그 결정된 데이터 셋 사이즈에 따라 부족한 값은 제로값을 추가, 즉, 제로 패딩(zero padding)을 적용한다. 이로써, 제어모듈(150)은 획득된 EMR 데이터에 대한 전처리를 수행 완료하도록 할 수 있다.

[0046] 그 외, 제어모듈(150)의 구체적인 동작에 대해서는 이하에서 각각의 도면을 기반으로 설명하도록 한다.

[0047] 도 2는 본 개시의 일 실시예에 따른 환자에 대한 위험도 및 악화 예측을 위한 데이터 처리 방법을 나타내는 순서도이다.

[0048] 도 2를 참조하면, 데이터 처리 장치(100)는 저장모듈(130) 또는 별도의 데이터베이스에서 적어도 하나의 환자 중 특정 환자자에 대한 EMR 데이터를 획득(수집)한다(S110).

[0049] 여기서, EMR 데이터는 의료진에 의하여 입력되는 데이터로서, 진료기록, 검사, 투약, 증상 및 처치 처방 중 적어도 하나에 대한 데이터를 포함할 수 있다. 이때, 검사에 대한 데이터는 활력징후(vital sign)에 대한 측정 데이터, 혈액검사에 따른 데이터 등 수치형 데이터를 포함할 수 있다. 구체적으로, 활력징후는 의료진이 환자의 상태를 수치화하여 판단하는 가장 기초적인 정보인 동시에, 의료진이 환자를 파악하려는 의도를 가지고 측정 및 기록하는 정보로서, 맥박, 호흡, 혈압, 체온으로 이루어질 수 있다.

[0050] 그 다음으로, 데이터 처리 장치(100)는 제로 패딩(zero padding)을 적용하여 S110 단계에 의해 획득된 EMR 데이터에 대한 전처리를 수행한다(S120).

[0051] 그 다음으로, 데이터 처리 장치(100)는 S120 단계에 의해 전처리된 EMR 데이터를 기반으로 기 설정된 인공지능 기반의 학습모델을 이용하여 그 특정 환자에 대한 위험도 및 악화 예측에 대한 결과정보를 생성한다(S130).

[0052] 앞서 설명한 바와 같이, 본 개시는 이상치가 제거된 시계열 데이터를 이용하도록 함으로써, 학습모델을 통해 도출할 수 있는 환자의 위험도 및 악화 예측에 대한 정확도를 향상시킬 수 있도록 한다. 이때, 환자의 위험도 및 악화 예측에 대한 정확도는 발생위험도 (risk) 또는 발생 가능성 (probability) 및 신뢰범위 (Confidence Interval)를 통해 제공될 수 있다.

[0053] EMR 데이터는 양이 많다고 하더라도 정규분포를 띄지 않는 경우나 한쪽으로 치우친(Skewed) 분포를 갖는 경우가 흔하다. 특히, 이상치에 의해 특정 분포를 잘 나타내지 못한다. 또한, 일반적으로 대부분 1^{st}-99^{th} 백분위수(percentile)로 데이터를 처리하고 있으나, 이는 반복측정데이터에서 정상치에 치중된 데이터 전처리가 되기 쉬운 문제점이 있다. 따라서, 이를 극복하고자 데이터 분포에서 추출한 백분위수의 정보에 임상적 지식을 이용하여 적절한 이상치 처리의 정의를 각 혈액정보 별로 구분하고, 그에 대한 분포를 임상적으로 검토하여 적절한 데이터 처리 여부를 확인할 수 있다.

[0054] 도 3은 본 개시의 일 실시예에 따라 EMR 데이터에 대한 전처리를 수행하는 동작을 구체적으로 나타내는 순서도이다.

[0055] 도 3을 참조하면, 데이터 처리 장치(100)는 S110 단계에 의해 획득된 EMR 데이터의 변수별 분포를 확인하고 이를 통해 EMR 데이터에 포함된 적어도 하나의 항목 중 특정 항목에 대한 이상치(Outlier)를 확인하고(S111), 그 확인된 이상치를 제거한다(S112). 이로써, 그 특정 항목에 대해 반복 측정된 시계열 데이터에 대한 이상치를 수정할 수 있다.

[0056] 그 다음으로, 데이터 처리 장치(100)는 기 설정된 기준정보를 기반으로 최적의 데이터 셋 사이즈를 결정하고(S113), S111 단계에 의해 이상치가 제거된 EMR 데이터를 이용하여 그 결정된 데이터 셋 사이즈에 따라 특정 항목에 대한 데이터 셋을 구축한다(S114). 여기서, 기 설정된 기준정보는 질병, 증상 및 시계열 자료상의 환자 상태에 따라 최적의 데이터셋 사이즈 및 참조값이 매핑되어 저장된 것일수 있다.

[0057] 한편, S114 단계에서, 그 결정된 데이터 셋 사이즈에 따라 부족한 값은 제로값을 추가, 즉, 제로 패딩(zero

padding)을 적용함으로써 데이터 셋을 구축할 수 있다.

[0058] 이러한 전처리 동작을 수행함으로써, EMR 데이터에 포함된 적어도 하나의 항목 중 특정 항목에 대한 이상치(Outlier)를 제거하는 동시에 최적의 사이즈를 갖는 데이터 셋을 구축함으로써, 길이가 길수록 예측력이 높아지는 시계열 데이터의 특성에 따라 한정된 데이터를 가지고도 높은 정확도를 갖는 예측 결과를 도출할 수 있다.

[0059] 도 4는 본 개시의 일 실시예에 따른 전처리 동작을 통해 EMR 데이터에 포함된 적어도 하나의 항목 중 키 항목에 대한 데이터 셋을 구축하는 일 예시를 나타내는 도면이다.

[0060] 먼저, 하기와 같은 두 가지 가정을 전제로 할 수 있다.

[0061] 1) 같은 환자의 키의 변화는 병원을 방문하는 동안 미미할 것이다.

[0062] 2) 같은 환자의 키 값을 그룹화하여 묶으면 중앙값(median), 차이값(difference)으로 오류치를 제거할 수 있을 것이다.

[0063] 이 중앙값 및 차이값을 계산할 때 이상치들이 오류로 작동하므로, 뚜렷한 에러값을 먼저 제거한다. 예를 들어, 뚜렷한 에러값은 100cm 이하, 200cm 이상으로 설정할 수 있다.

[0064] 이후, 환자별/날짜별로 데이터를 정렬하고, 같은 환자 내에서 키의 변화를 정방향(diff) 또는 역방향(rdiff)로 확인할 수 있다. 같은 환자 내에서 키 입력 날짜 차이를 계산하되, 짧은 기간의 차이는 오류로 간주할 수 있다.

[0065] 또한, 예를 들어, 측정된 키의 EMR 상 기록된 상한치와 하한치의 차이가 1cm 이상이고, 해당 차이가 한달 간격으로 했을 때 5cm 이상인 경우, 위아래 차이가 10cm 이상인 경우 및 중앙값에서 5cm 이상 차이나는 경우 제거하도록 한다. 즉, 중앙값 및 차이값을 각각에 대해 기 설정된 임계치와 비교하고, 그 설정된 임계치를 벗어나는 값은 제거하는 동작을 수행한다.

[0066] 도 5는 본 개시의 일 실시예에 따른 전처리 동작을 통해 EMR 데이터에 포함된 적어도 하나의 항목 중 체중 항목에 대한 데이터 셋을 구축하는 일 예시를 나타내는 도면이다.

[0067] 일 예로서, 각 컬럼(column) 별로 이상치를 0.1%, 99.9%로 계산하고, 해당 컬럼 바깥에 있는 밸류(value)값들이 실제로 이상한지를 평가한다.

[0068] 만약, 체중이 25kg 이하 또는 BMI가 8 이하인 경우, 및 체중이 120kg 이상이면서 BMI가 60 이상이거나 BMI가 없는 경우, 이를 뚜렷한 이상치로 판단하여 제거한다.

[0069] 또한, 예를 들어, 체중 전후 차이가 10kg 이상인 경우, 체중 전/후(또는 조건) 차이가 80kg 이상이면서 100kg 이상인 경우 제거하도록 한다.

[0070] 앞서, 도 4 및 도 5를 통해 살펴본 바와 같이, 시계열 데이터를 통해 그 특정 항목에 대한 값들이 연속성을 가지는지 여부를 판단하고, 그 연속성에서 벗어나는 값을 이상치로 간주하여 제거하는 것이다. 이때, 연속성은 과거 기록값, 기록값의 변화폭을 통해 판단할 수 있으며, 예를 들어, 그 변화폭이 기 설정된 임계치의 범위를 벗어나면 이상치로 간주할 수 있다.

[0071] 한편, 그 외에도 EMR 데이터에서 키 및 체중을 반복 측정한 시계열 데이터를 이용하는 경우, 이상치를 찾되, 여러 개의 이상치가 겹친 경우에는 이를 에러값으로 간주한다. 이때, 이상치를 찾되, 1회 측정일 경우 평가불가, 2회 측정에서 서로 값이 상이한 경우 정상범위에서 먼 값을 이상치로 간주하며, 2회 측정에서 동일한 이상치가 겹친 경우에는 이를 정상값으로 간주한다. 성인의 경우 1개월 이내 간격으로 연속 측정한 값이 15%초과할 경우 정상치에서 먼 값을 이상치로 간주하며 소아의 경우 연령에 따라 보정한다.

[0072] 한편, 전처리를 수행하기 이전에 결측치가 발생하는지 여부를 판단하여, 그 획득된 EMR 데이터가 위험도 및 악화를 예측하기에 용이한(적절한) 환자 데이터인지 여부를 판단하도록 할 수 있다. 또한, 시계열 반복측정 데이터의 특정 구간(윈도)에서 결측치가 5%미만일 경우 해당구간의 결측치는 이동평균을 이용한 보간법 등의 방법으로 채울 수 있다

[0073] 예를 들어, 환자의 입원 별로 시계열 상에서 생성되는 활력징후에 대한 데이터의 결측치가 발생하는 것을 전체 입원일수에 따라 분율로 평가함으로써, 적절한 환자 데이터인지 여부를 판단할 수 있다. 만약, 활력징후에 대한 데이터가 50% 이상 누락되었거나, 실험실 데이터가 80% 이상 누락되었거나, 전체 입원 기간 동안 특정 질병에 대한 중증도 결과 데이터가 누락된 환자는 제거하도록 할 수 있다.

[0074] 도 6은 본 개시의 일 실시예에 따른 기 설정된 인공지능 기반의 학습모델을 이용하여 서로 다른 데이터 셋 사이

즈를 갖는 데이터 셋을 기반으로 예측된 결과의 일 예시를 나타내는 도면이다.

[0075] 도 6을 참조하면, (A)의 모델에는 입원기간 2일에 대해 구축된 데이터 셋을 입력하고, (B)의 모델에는 입원기간 10일에 대해 구축된 데이터 셋을 입력한 것이다. 즉, (A)의 모델 및 (B)의 모델의 윈도우 크기는 서로 상이하다.

[0076] 구체적으로, 입력 윈도우가 2일인 (A)의 모델은 단지 첫번째 시간에 대해서만 제로 패딩이 적용되고, 입력 윈도우가 10일인 (B)의 모델은 10일의 입원기간 동안까지 제로 패딩이 적용된다. 즉, (B)의 모델을 통해 더 많은 정보를 도출할 수 있다. 이 패딩 윈도우는 마스킹 값에 의해 귀속되었으나, 입력 크기의 길이와 시간 임베딩 추가 측면에서 두 모델 간 차이가 있다.

[0077] 여기서, D0은 마지막 입력 윈도우 날짜의 결과 이벤트를 나타내고, D1은 마지막 입력 윈도우 날짜로부터 1일 후의 결과 이벤트를 나타내며, D2는 마지막 입력 윈도우 날짜로부터 2일 후의 결과 이벤트를 나타낸다.

[0078] 도 7은 본 개시의 일 실시예에 따른 인공지능 기반의 학습모델의 기본 구조를 나타내는 도면이다.

[0079] 도 7을 참조하면, 입력값은 시계열 데이터의 발생 주기에 따라 구분되어 사용되며, 전체 1회 발생 (키, 나이 등의 demographic data), 1일 1회 발생 (혈액검사 소견 등), 1일 3회 발생 (혈압 맥박 등 혈역학 변수) 로 구분하여 각각의 트랜스포머 블록으로 입력된다. 심층신경망 레이어에서 각각의 트랜스포머에서 전달된 일부가 사용될 수 있으며 각각 독립적으로 디코더로 전달되어 결과값을 산출한다.

[0080] 구체적으로, 인공지능 기반의 학습모델은 입력 데이터 셋을 확인하는 빈도에 따라 설정된 두 부분으로 구성할 수 있다. 활력징후(Vital)는 매 8시간마다, 환자가 불안정할 때 더 자주 확인되었고, 실험실(Lab) 데이터는 보통 24시간마다 기록될 수 있다. 만약, 활력징후가 실험실 데이터 빈도(1일)를 따르는 경우, 데이터의 추세 기능이 손실된다. 한편, 실험실 데이터가 활력징후 데이터 빈도(8시간)를 따르는 경우, 증가된 데이터 빈도는 이전 값을 사용하여 실험실 데이터를 채워야 한다.

[0081] 그러나, 이는 하나의 실시예일 뿐, 입력 데이터 셋이 더 추가되어 구성될 수 있으며, 활력징후나 실험실 데이터 외에도 증상에 대한 기록에 따른 입력 데이터 셋을 더 이용할 수 있다.

[0082] 앞서 설명한 본 개시에 따른 인공지능 기반의 학습모델로서 랜덤 포레스트 모델, XG 부스트 모델, 심층 신경망 모델 및 로지스틱 회귀 모델 중 적어도 하나를 적용하여 이용할 수 있다.

[0083] 도 8은 본 개시에 따른 인공지능 기반의 학습모델에 의해 생성된 환자에 대한 위험도 및 악화 예측에 대한 결과 정보의 일 예시를 나타내는 도면으로서, 입원기간 중 일일 위험도의 예측을 나타낸 것이다.

[0084] 도 8을 참조하면, 목표는 실제 일일 위험도이고, 인공지능 기반의 학습모델에 의한 0~2일까지의 예측에 대한 결과이다. 해당 환자는 2일차에 중증 상태로 악화되었고, 9일차에 회복되어, 13일차에 경도 상태를 보인다. 이 학습모델은 악화 및 회복 전 2일 동안에 대해 예측을 수행한다.

[0085] 한편, 앞서 설명한 본 개시에 따른 환자 위험도 및 악화 예측을 위한 데이터 처리는 각종 질병에 적용될 수 있는데, 예를 들어, 그 질병의 종류는 주로는 중환자실 내에서 코로나 질환의 악화, 산소호흡기 이탈여부, 심혈관계 합병증 및 급성 신손상 등이 될 수 있으며, 이를 한정하지는 않는다.

[0086] 상기 전술한 프로그램은, 상기 컴퓨터가 프로그램을 읽어 들여 프로그램으로 구현된 상기 방법들을 실행시키기 위하여, 상기 컴퓨터의 프로세서(CPU)가 상기 컴퓨터의 장치 인터페이스를 통해 읽힐 수 있는 C, C++, JAVA, 기계어 등의 컴퓨터 언어로 코드화된 코드(Code)를 포함할 수 있다. 이러한 코드는 상기 방법들을 실행하는 필요한 기능들을 정의한 함수 등과 관련된 기능적인 코드(Functional Code)를 포함할 수 있고, 상기 기능들을 상기 컴퓨터의 프로세서가 소정의 절차대로 실행시키는데 필요한 실행 절차 관련 제어 코드를 포함할 수 있다. 또한, 이러한 코드는 상기 기능들을 상기 컴퓨터의 프로세서가 실행시키는데 필요한 추가 정보나 미디어가 상기 컴퓨터의 내부 또는 외부 메모리의 어느 위치(주소 번지)에서 참조되어야 하는지에 대한 메모리 참조관련 코드를 더 포함할 수 있다. 또한, 상기 컴퓨터의 프로세서가 상기 기능들을 실행시키기 위하여 원격(Remote)에 있는 어떠한 다른 컴퓨터나 서버 등과 통신이 필요한 경우, 코드는 상기 컴퓨터의 통신 모듈을 이용하여 원격에 있는 어떠한 다른 컴퓨터나 서버 등과 어떻게 통신해야 하는지, 통신 시 어떠한 정보나 미디어를 송수신해야 하는지 등에 대한 통신 관련 코드를 더 포함할 수 있다.

[0087] 상기 저장되는 매체는, 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상기 저

장되는 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있지만, 이에 제한되지 않는다. 즉, 상기 프로그램은 상기 컴퓨터가 접속할 수 있는 다양한 서버 상의 다양한 기록매체 또는 사용자의 상기 컴퓨터상의 다양한 기록매체에 저장될 수 있다. 또한, 상기 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장될 수 있다.

[0088] 본 개시의 실시예와 관련하여 설명된 방법 또는 알고리즘의 단계들은 하드웨어로 직접 구현되거나, 하드웨어에 의해 실행되는 소프트웨어 모듈로 구현되거나, 또는 이들의 결합에 의해 구현될 수 있다. 소프트웨어 모듈은 RAM(Random Access Memory), ROM(Read Only Memory), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM), 플래시 메모리(Flash Memory), 하드 디스크, 착탈형 디스크, CD-ROM, 또는 본 개시가 속하는 기술 분야에서 잘 알려진 임의의 형태의 컴퓨터 판독가능 기록매체에 상주할 수도 있다.

[0089] 이상, 첨부된 도면을 참조로 하여 본 개시의 실시예를 설명하였지만, 본 개시가 속하는 기술분야의 통상의 기술자는 본 개시가 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로, 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며, 제한적이 아닌 것으로 이해해야만 한다.

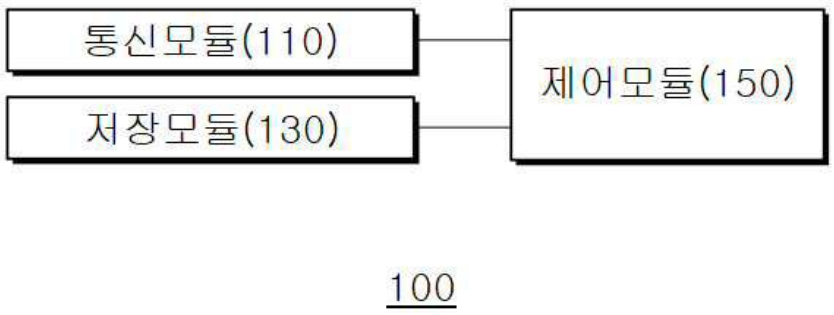
부호의 설명

[0090] 100 : 데이터 처리 장치 110 : 통신모듈

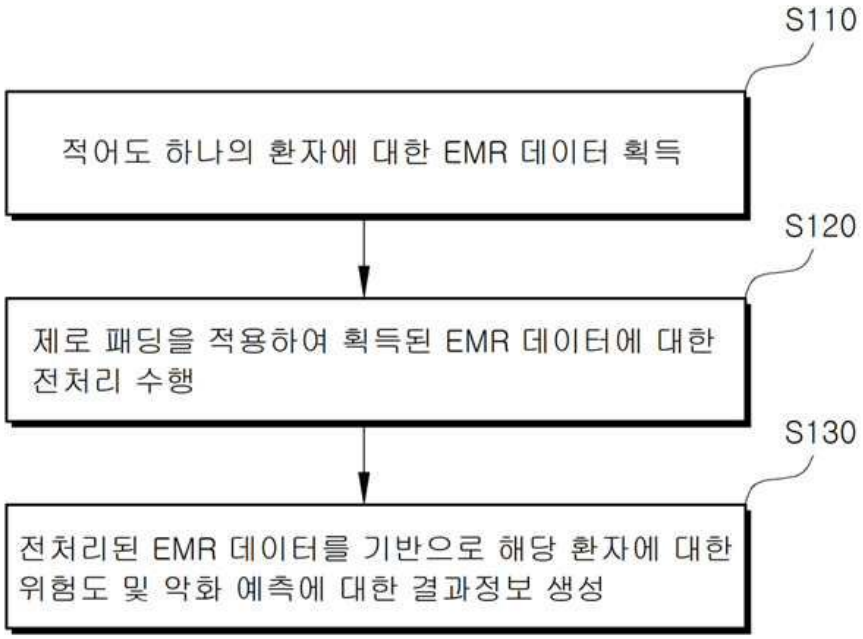
130 : 저장모듈 150 : 제어모듈

도면

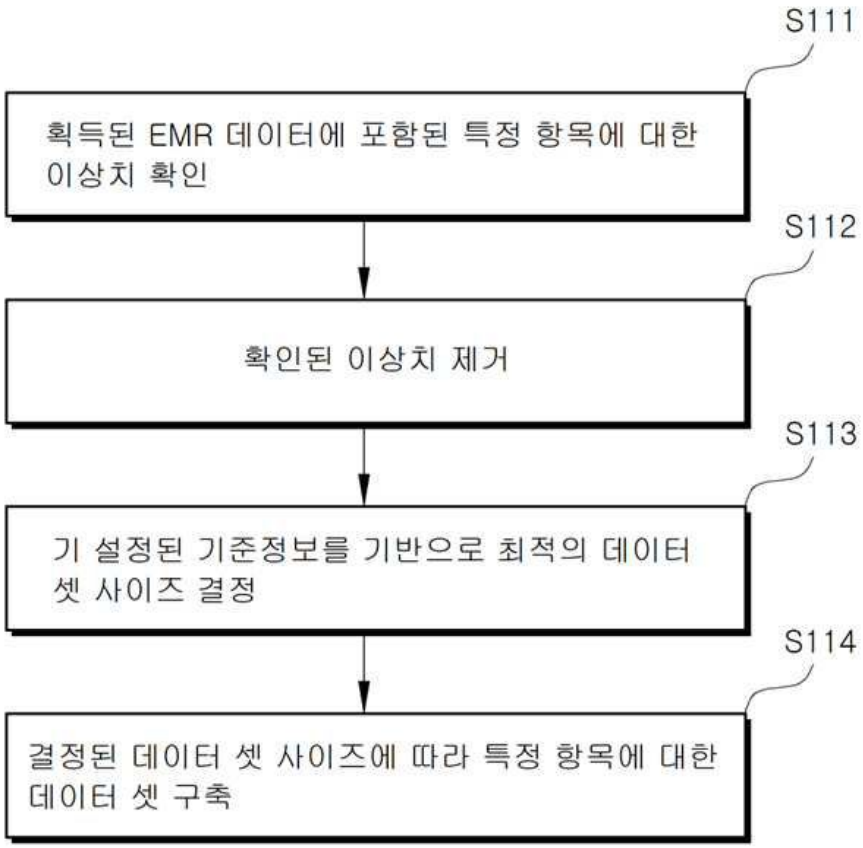
도면1



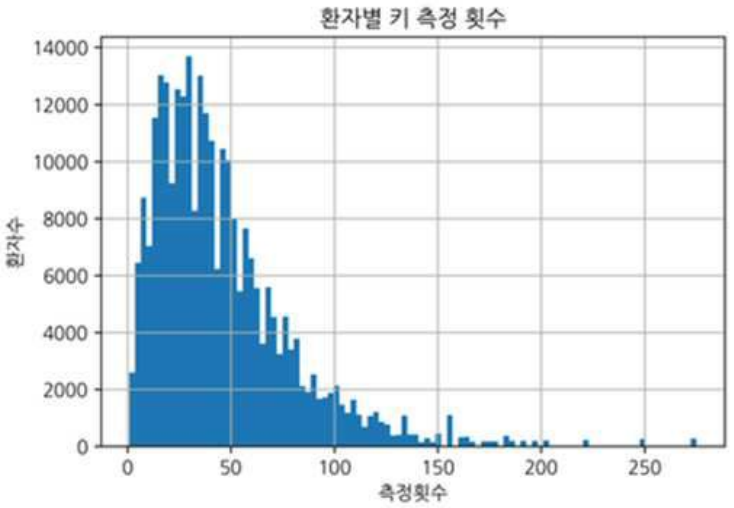
도면2


S110
적어도 하나의 환자에 대한 EMR 데이터 획득
S120
제로 패딩을 적용하여 획득된 EMR 데이터에 대한 전처리 수행
S130
전처리된 EMR 데이터를 기반으로 해당 환자에 대한 위험도 및 악화 예측에 대한 결과정보 생성

도면3

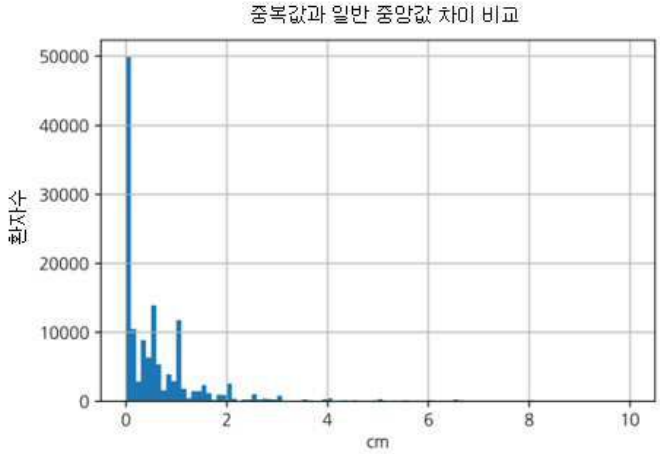

S111
획득된 EMR 데이터에 포함된 특정 항목에 대한 이상치 확인
S112
확인된 이상치 제거
S113
기 설정된 기준정보를 기반으로 최적의 데이터 셋 사이즈 결정
S114
결정된 데이터 셋 사이즈에 따라 특정 항목에 대한 데이터 셋 구축

도면4



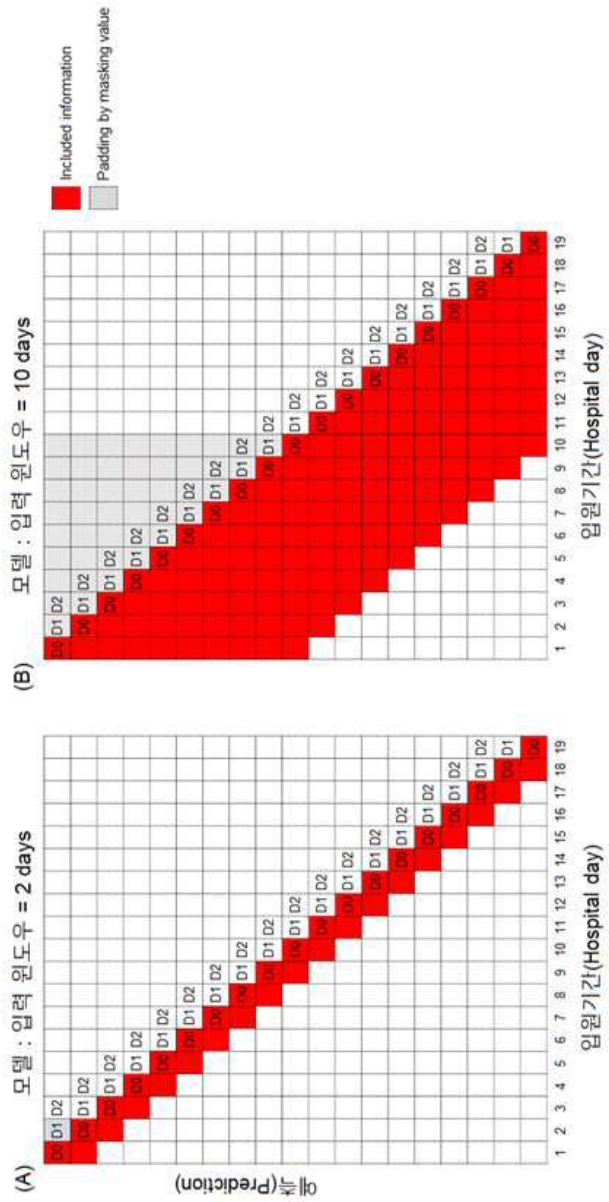
	연구번호	height	weight	bmi	count	median	diff	outlier
0.001	1.167993e+09	52.0	29.25	11.606979	1.0	137.00	0.0000	1.0
0.010	1.168495e+09	141.2	35.00	14.219959	4.0	143.00	0.0000	1.0
0.100	1.174991e+09	150.9	45.00	17.735754	12.0	151.00	0.0000	1.0
0.900	1.208849e+09	173.5	74.20	27.181630	89.0	173.40	1.6000	1.0
0.990	1.213678e+09	181.2	93.60	34.824288	164.0	181.00	7.6000	1.0
0.999	1.215619e+09	186.9	158.00	180.343221	275.0	186.75	110.1028	1.0

도면5

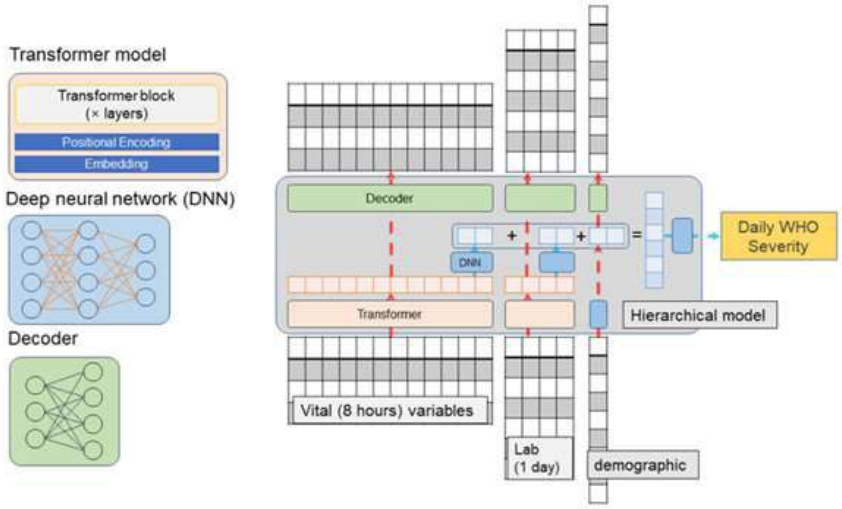


	연구번호	접수일자	height	month_diff	height_diff	height_rdiff	height_d_diff
902	1168116176	2010-01-08	170.0	0.000000	7.0	0.0	inf
903	1168116176	2010-05-26	170.0	4.600000	0.0	0.0	0.000000
904	1168116176	2010-07-01	170.0	1.200000	0.0	28.5	0.000000
905	1168116176	2010-08-06	141.5	1.200000	28.5	28.5	23.750000
906	1168116176	2010-08-13	170.0	0.233333	28.5	2.0	122.142857
907	1168116176	2011-05-08	168.0	8.933333	2.0	0.0	0.223881
908	1168116176	2011-05-24	168.0	0.533333	0.0	6.0	0.000000
909	1168116176	2011-11-07	162.0	5.566667	6.0	3.0	1.077844
910	1168116176	2012-02-08	165.0	3.100000	3.0	NaN	0.967742

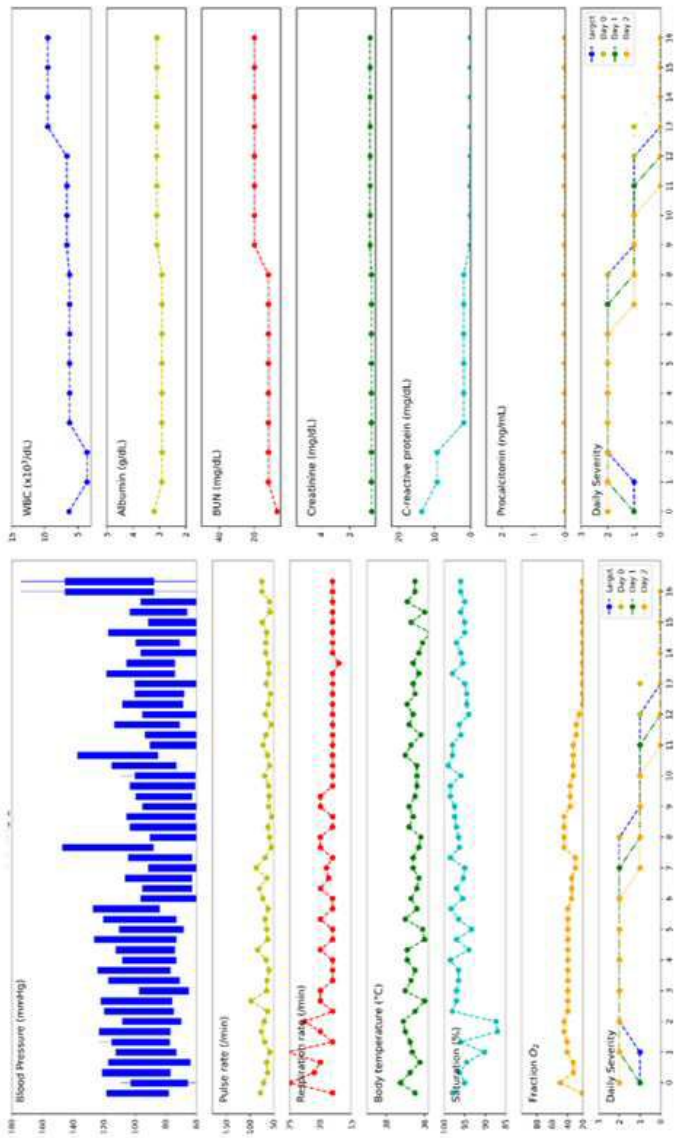
도면6


Included information
Padding by masking value
(A)
모델 : 입력 윈도우 = 2 days
(B)
모델 : 입력 윈도우 = 10 days
예측(Prediction)
입원기간(Hospital day)
입원기간(Hospital day)

도면7


Transformer model
Transformer block
(× layers)
Positional Encoding
Embedding
Deep neural network (DNN)
Decoder
Decoder
DNN
Transformer
Daily WHO
Severity
Hierarchical model
Vital (8 hours) variables
Lab
(1 day)
demographic

도면8


WBC (x10³/dL)
Albumin (g/dL)
BUN (mg/dL)
Creatinine (mg/dL)
C-reactive protein (mg/dL)
Procalcitonin (ng/mL)
Daily Severity
Blood Pressure (mmHg)
Pulse rate (/min)
Respiratory rate (/min)
Body temperature (°C)
Saturation (%)
Fraction O_2
Daily Severity