



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0101367  
(43) 공개일자 2024년07월02일

(51) 국제특허분류(Int. Cl.)

G16H 50/20 (2018.01) A61B 5/08 (2006.01)  
G06N 20/00 (2019.01) G06N 3/0442 (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 50/20 (2018.01)  
A61B 5/08 (2013.01)

(21) 출원번호 10-2023-0161207

(22) 출원일자 2023년11월20일

심사청구일자 2023년11월20일

(30) 우선권주장

1020220182591 2022년12월23일 대한민국(KR)

(71) 출원인

순천향대학교 산학협력단

충청남도 아산시 신창면 순천향로 22, 순천향대학교내

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

길효욱

충청남도 천안시 서북구 시청로 73, 311동 504호  
(불당동, 불당마을동일하이빌아파트)

조남준

충청남도 천안시 동남구 순천향6길 31 (봉명동)  
(뒷면에 계속)

(74) 대리인

특허법인 무한

전체 청구항 수 : 총 15 항

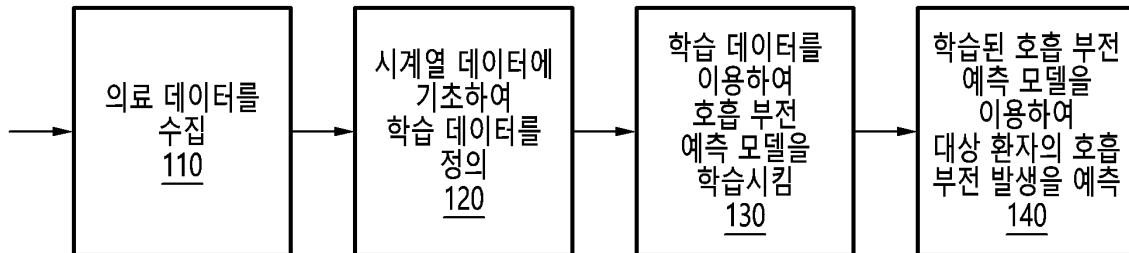
(54) 발명의 명칭 호흡 부전 예측 모델을 이용하여 농약 중독 환자의 호흡 부전 발생을 예측하는 예측 장치 및 예측 방법

(57) 요약

호흡 부전 예측 모델을 이용하여 농약 중독 환자의 호흡 부전 발생을 예측하는 예측 장치 및 예측 방법이 개시된다. 예측 장치는 실행 가능한 명령어들을 저장하는 메모리 및 프로세서를 포함하고, 프로세서에 의해 실행 가능한 명령어가 실행되는 경우, 예측 장치는, 제1 농약 중독 환자의 환자 정보 데이터, 제1 농약 중독 환자의 생체

(뒷면에 계속)

대표도 - 도1



징후 정보를 포함하는 생체 징후 데이터 및 제1 농약 중독 환자에 대한 임상병리 검사 결과 정보를 포함하는 랩 코드 데이터를 포함하는 제1 농약 중독 환자의 의료 데이터를 수신하고, 수신한 의료 데이터에서 결측 값을 보간하여 결측 값이 보간된 의료 데이터를 생성하고, 미리 정의된 시간 단위의 슬라이딩 윈도우에 기반하여 결측 값이 보간된 의료 데이터로부터 복수의 시계열 데이터를 획득하고, 획득된 복수의 시계열 데이터를 기초로 농약 중독 환자의 호흡 부전 발생을 예측하는 호흡 부전 예측 모델을 학습시키고, 호흡 부전 발생 예측의 대상이 되는 제2 농약 중독 환자의 의료 데이터를 수신하고, 제2 농약 중독 환자의 의료 데이터에 대해 데이터 전처리를 수행하고, 학습된 호흡 부전 예측 모델에 데이터 전처리가 수행된 제2 농약 중독 환자의 의료 데이터를 입력하는 것에 의해 제2 농약 중독 환자의 호흡 부전 발생과 관련된 예측 데이터를 획득할 수 있다.

(52) CPC특허분류

G06N 20/00 (2021.08)

G06N 3/0442 (2023.01)

G06N 3/08 (2023.01)

G16H 10/60 (2021.08)

G16H 50/50 (2018.01)

G16H 50/70 (2018.01)

(72) 발명자

이화민

충청남도 아산시 배방읍 연화로 11, 607동 902호  
(연화마을STXKAN6단지아파트)

김영민

충청남도 천안시 서북구 서부19길 56, 403호 (성정동)

이 발명을 지원한 국가연구개발사업

|             |                           |
|-------------|---------------------------|
| 과제고유번호      | 1711157181                |
| 과제번호        | 2019M3E5D1A02069073       |
| 부처명         | 과학기술정보통신부                 |
| 과제관리(전문)기관명 | 한국연구재단                    |
| 연구사업명       | 과학기술혁신인재양성                |
| 연구과제명       | 일회용 패치 기반의 환자 모니터링 시스템 개발 |
| 기 여 율       | 75/100                    |
| 과제수행기관명     | 순천향대학교                    |
| 연구기간        | 2022.01.01 ~ 2022.12.31   |

이 발명을 지원한 국가연구개발사업

|             |                         |
|-------------|-------------------------|
| 과제고유번호      | 1711174057              |
| 과제번호        | 00156439                |
| 부처명         | 과학기술정보통신부               |
| 과제관리(전문)기관명 | 정보통신기획평가원               |
| 연구사업명       | 정보통신방송혁신인재양성(R&D)       |
| 연구과제명       | ICT혁신인재4.0(고려대학교)       |
| 기 여 율       | 25/100                  |
| 과제수행기관명     | 고려대학교산학협력단              |
| 연구기간        | 2022.07.01 ~ 2022.12.31 |

## 명세서

### 청구범위

#### 청구항 1

호흡 부전 예측 모델을 이용하여 농약 중독 환자의 호흡 부전 발생을 예측하는 예측 장치에 있어서,  
실행 가능한 명령어들을 저장하는 메모리; 및  
상기 프로세서를 포함하고,  
상기 프로세서에 의해 상기 실행 가능한 명령어가 실행되는 경우, 상기 예측 장치는,  
제1 농약 중독 환자의 환자 정보 데이터, 상기 제1 농약 중독 환자의 생체 징후 정보를 포함하는 생체 징후 데이터 및 상기 제1 농약 중독 환자에 대한 임상병리 검사 결과 정보를 포함하는 랩 코드 데이터를 포함하는 상기 제1 농약 중독 환자의 의료 데이터를 수신하고,  
상기 수신한 의료 데이터에서 결측 값을 보간하여 결측 값이 보간된 의료 데이터를 생성하고,  
미리 정의된 시간 단위의 슬라이딩 윈도우에 기반하여 상기 결측 값이 보간된 의료 데이터로부터 복수의 시계열 데이터를 획득하고,  
상기 획득된 복수의 시계열 데이터를 기초로 농약 중독 환자의 호흡 부전 발생을 예측하는 호흡 부전 예측 모델을 학습시키고,  
호흡 부전 발생 예측의 대상이 되는 제2 농약 중독 환자의 의료 데이터를 수신하고,  
상기 제2 농약 중독 환자의 의료 데이터에 대해 데이터 전처리를 수행하고,  
상기 학습된 호흡 부전 예측 모델에 상기 데이터 전처리가 수행된 상기 제2 농약 중독 환자의 의료 데이터를 입력하는 것에 의해 상기 제2 농약 중독 환자의 호흡 부전 발생과 관련된 예측 데이터를 획득하는,  
예측 장치.

#### 청구항 2

제1항에 있어서,  
상기 예측 장치는,  
상기 수신된 의료 데이터의 시계열 특성을 기초로 결측 값이 포함된 항목의 가장 최근에 측정된 값, 상기 수신한 의료 데이터에서 결측 값이 포함된 항목과 연관성이 가장 높은 항목의 측정 값 및 결측 값이 포함된 항목의 측정 값들의 평균 값 중 어느 하나를 기초로 결측 값을 보간하는,  
예측 장치.

#### 청구항 3

제1항에 있어서,  
상기 예측 장치는,  
상기 제1 농약 중독 환자의 하루 단위의 생체 징후 데이터에 미리 정의된 시간 단위의 슬라이딩 윈도우를 시간의 흐름에 따라 순차적으로 적용하여 상기 제1 농약 중독 환자에 대한 복수의 시계열 데이터를 획득하는,  
예측 장치.

#### 청구항 4

제1항에 있어서,

상기 획득된 복수의 시계열 데이터에 포함된 항목들은,

음독 농약, 음독량, 성별, 나이, 의식 상태, 혈압, 맥박, 체온, 백혈구 수치, 혈소판 수치, 알부민 수치, 총 이산화탄소(total CO<sub>2</sub>; TC02) 수치, C-반응 단백질(C-reactive protein) 수치, 산도, 산소 분압, 산소 포화도 및 젖산 수치를 포함하는,

예측 장치.

#### 청구항 5

제1항에 있어서,

상기 예측 장치는,

상기 획득된 시계열 데이터에 포함된 각 항목에 대한 최대 값 및 최소 값을 기초로 각 항목의 값들을 스케일링 하는,

예측 장치.

#### 청구항 6

제1항에 있어서,

상기 예측 장치는,

상기 획득된 복수의 시계열 데이터에 ADASYN(adaptive synthetic; ADASYN) 오버샘플링 기법을 적용하여 생성된 학습 데이터 세트를 기초로 상기 호흡 부전 예측 모델을 학습시키는,

예측 장치.

#### 청구항 7

제1항에 있어서,

상기 호흡 부전 예측 모델은,

로지스틱 회귀 모델, 결정 트리 모델, 랜덤 포레스트 모델 및 LSTM(long short-term memory; LSTM) 모델 중 적어도 하나를 포함하는 것인,

예측 장치.

#### 청구항 8

호흡 부전 예측 모델을 이용하여 농약 중독 환자의 호흡 부전 발생을 예측하는 예측 방법에 있어서,

제1 농약 중독 환자의 환자 정보 데이터, 상기 제1 농약 중독 환자의 생체 징후 정보를 포함하는 생체 징후 데이터 및 상기 제1 농약 중독 환자에 대한 임상병리 검사 결과 정보를 포함하는 랩 코드 데이터를 포함하는 상기 제1 농약 중독 환자의 의료 데이터를 수신하는 동작;

상기 수신한 의료 데이터에서 결측 값을 보간하여 결측 값이 보간된 의료 데이터를 생성하는 동작;

미리 정의된 시간 단위의 슬라이딩 윈도우에 기반하여 상기 결측 값이 보간된 의료 데이터로부터 복수의 시계열 데이터를 획득하는 동작;

상기 획득된 복수의 시계열 데이터를 기초로 농약 중독 환자의 호흡 부전 발생을 예측하는 호흡 부전 예측 모델을 학습시키는 동작;

호흡 부전 발생 예측의 대상이 되는 제2 농약 중독 환자의 의료 데이터를 수신하는 동작;

상기 제2 농약 중독 환자의 의료 데이터에 대해 데이터 전처리를 수행하는 동작; 및

상기 학습된 호흡 부전 예측 모델에 상기 데이터 전처리가 수행된 상기 제2 농약 중독 환자의 의료 데이터를 입력하는 것에 의해 상기 제2 농약 중독 환자의 호흡 부전 발생과 관련된 예측 데이터를 획득하는 동작

을 포함하는,

예측 방법.

## 청구항 9

제8항에 있어서,

상기 보관된 의료 데이터를 생성하는 동작은,

상기 수신된 의료 데이터의 시계열 특성을 기초로 결측 값이 포함된 항목의 가장 최근에 측정된 값, 상기 수신한 의료 데이터에서 결측 값이 포함된 항목과 연관성이 가장 높은 항목의 측정 값 및 결측 값이 포함된 항목의 측정 값들의 평균 값 중 어느 하나를 기초로 결측 값을 보관하는 동작을 포함하는,

예측 방법.

## 청구항 10

제8항에 있어서,

상기 시계열 데이터를 획득하는 동작은,

상기 제1 농약 중독 환자의 하루 단위의 생체 징후 데이터에 미리 정의된 시간 단위의 슬라이딩 윈도우를 시간의 흐름에 따라 순차적으로 적용하여 상기 제1 농약 중독 환자에 대한 복수의 시계열 데이터를 획득하는 동작을 포함하는,

예측 방법.

## 청구항 11

제8항에 있어서,

상기 획득된 복수의 시계열 데이터에 포함된 항목들은,

음독 농약, 음독량, 성별, 나이, 의식 상태, 혈압, 맥박, 체온, 백혈구 수치, 혈소판 수치, 알부민 수치, 총 이산화탄소(total CO<sub>2</sub>; TC02) 수치, C-반응 단백질(C-reactive protein) 수치, 산도, 산소 분압, 산소 포화도 및 젖산 수치를 포함하는,

예측 방법.

## 청구항 12

제8항에 있어서,

상기 호흡 부전 발생을 예측하는 예측 방법에 있어서,

상기 획득된 시계열 데이터에 포함된 각 항목에 대한 최대 값 및 최소 값을 기초로 각 항목의 값들을 스케일링

하는 동작을 더 포함하는,  
예측 방법.

### 청구항 13

제8항에 있어서,  
호흡 부전 예측 모델을 학습시키는 동작은,  
상기 획득된 복수의 시계열 데이터에 ADASYN(adaptive synthetic; ADASYN) 오버샘플링 기법을 적용하여 생성된 학습 데이터 세트를 기초로 상기 호흡 부전 예측 모델을 학습시키는 동작을 포함하는,  
예측 방법.

### 청구항 14

제8항에 있어서,  
상기 호흡 부전 예측 모델은,  
로지스틱 회귀모델, 결정 트리 모델, 랜덤 포레스트 모델 및 LSTM(long short-term memory; LSTM) 모델 중 적어도 하나를 포함하는 것인,  
예측 방법.

### 청구항 15

하드웨어와 결합되어 제8항 내지 제14항 중 어느 하나의 항의 방법을 실행시키기 위하여 컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램.

## 발명의 설명

### 기술 분야

[0001] 본 개시는 호흡 부전 예측 모델을 이용하여 농약 중독 환자의 호흡 부전 발생을 예측하는 예측 장치 및 예측 방법에 관한 것이다.

### 배경 기술

[0002] 일반적으로 농약 중독 환자에 대한 호흡 부전은 생징후, 혈액 검사 데이터, 소변 검사 데이터 및 임상경험을 기초로 예측될 수 있다. 호흡 부전 예측을 위해 농약 중독 환자에 대한 위험도를 측정하는 중독 중증도 지수(poisoning severity score)는 환자의 수가 적고, 중독 중증도 지수에 대한 내용이 업데이트되지 않아 현재 사용되지 않고 있다. 단일 물질로 이루어진 농약에 관한 연구가 보고되나, 임상 실험이 시행되지 않았고, 농약 중독을 일으키는 농약이 확인되지 않은 경우 호흡 부전 예측을 위해 사용될 수 없다.

## 발명의 내용

### 해결하려는 과제

### 과제의 해결 수단

[0003] 일 실시예에 따른 호흡 부전 예측 모델을 이용하여 농약 중독 환자의 호흡 부전 발생을 예측하는 예측 장치는,

실행 가능한 명령어들을 저장하는 메모리 및 상기 프로세서를 포함하고, 상기 프로세서에 의해 상기 실행 가능한 명령어가 실행되는 경우, 상기 예측 장치는, 제1 농약 중독 환자의 환자 정보 데이터, 상기 제1 농약 중독 환자의 생체 징후 정보를 포함하는 생체 징후 데이터 및 상기 제1 농약 중독 환자에 대한 임상병리 검사 결과 정보인 랩 코드 데이터를 포함하는 상기 제1 농약 중독 환자의 의료 데이터를 수신하고, 상기 수신한 의료 데이터에서 결측 값을 보간하여 결측 값이 보간된 의료 데이터를 생성하고, 미리 정의된 시간 단위의 슬라이딩 윈도우에 기반하여 상기 결측 값이 보간된 의료 데이터로부터 복수의 시계열 데이터를 획득하고, 상기 획득된 복수의 시계열 데이터를 기초로 농약 중독 환자의 호흡 부전 발생을 예측하는 호흡 부전 예측 모델을 학습시키고, 호흡 부전 발생 예측의 대상이 되는 제2 농약 중독 환자의 의료 데이터를 수신하고, 상기 제2 농약 중독 환자의 의료 데이터에 대해 데이터 전처리를 수행하고, 상기 학습된 호흡 부전 예측 모델에 상기 데이터 전처리가 수행된 상기 제2 농약 중독 환자의 의료 데이터를 입력하는 것에 의해 상기 제2 농약 중독 환자의 호흡 부전 발생과 관련된 예측 데이터를 획득할 수 있다.

[0004] 상기 예측 장치는, 상기 수신된 의료 데이터의 시계열 특성을 기초로 결측 값이 포함된 항목의 가장 최근에 측정된 값, 상기 수신한 의료 데이터에서 결측 값이 포함된 항목과 연관성이 가장 높은 항목의 측정 값 및 결측 값이 포함된 항목의 측정 값들의 평균 값 중 어느 하나를 기초로 결측 값을 보간할 수 있다.

[0005] 상기 예측 장치는, 상기 제1 농약 중독 환자의 하루 단위의 생체 징후 데이터에 미리 정의된 시간 단위의 슬라이딩 윈도우를 시간의 흐름에 따라 순차적으로 적용하여 상기 제1 농약 중독 환자에 대한 복수의 시계열 데이터를 획득할 수 있다.

[0006] 상기 획득된 복수의 시계열 데이터에 포함된 항목들은, 음독 농약, 음독량, 성별, 나이, 의식 상태, 혈압, 맥박, 체온, 백혈구 수치, 혈소판 수치, 알부민 수치, 총 이산화탄소(total CO<sub>2</sub>; TC0<sub>2</sub>) 수치, C-반응 단백질(C-reactive protein) 수치, 산도, 산소 분압, 산소 포화도 및 젖산 수치를 포함할 수 있다.

[0007] 상기 예측 장치는, 상기 획득된 시계열 데이터에 포함된 각 항목에 대한 최대 값 및 최소 값을 기초로 각 항목의 값들을 스케일링할 수 있다.

[0008] 상기 예측 장치는, 상기 획득된 복수의 시계열 데이터에 ADASYN(adaptive synthetic; ADASYN) 오버샘플링 기법을 적용하여 생성된 학습 데이터 세트를 기초로 상기 호흡 부전 예측 모델을 학습시킬 수 있다.

[0009] 상기 호흡 부전 예측 모델은, 로지스틱 회귀 모델, 결정 트리 모델, 랜덤 포레스트 모델 및 LSTM(long short-term memory; LSTM) 모델 중 적어도 하나를 포함하는 것일 수 있다.

[0010] 일 실시예에 따른, 호흡 부전 예측 모델을 이용하여 농약 중독 환자의 호흡 부전 발생을 예측하는 예측 방법은 제1 농약 중독 환자의 환자 정보 데이터, 상기 제1 농약 중독 환자의 생체 징후 정보를 포함하는 생체 징후 데이터 및 상기 제1 농약 중독 환자에 대한 임상병리 검사 결과 정보를 포함하는 랩 코드 데이터를 포함하는 상기 제1 농약 중독 환자의 의료 데이터를 수신하는 동작, 상기 수신한 의료 데이터에서 결측 값을 보간하여 결측 값이 보간된 의료 데이터를 생성하는 동작, 미리 정의된 시간 단위의 슬라이딩 윈도우에 기반하여 상기 결측 값이 보간된 의료 데이터로부터 복수의 시계열 데이터를 획득하는 동작, 상기 획득된 시계열 데이터에 포함된 각 항목에 대한 최대 값 및 최소 값을 기초로 스케일링하는 동작, 상기 복수의 시계열에 오버샘플링 기법을 적용하여 데이터 불균형을 해결하는 동작, 상기 획득된 시계열 데이터를 기초로 농약 중독 환자의 호흡 부전 발생을 예측하는 호흡 부전 예측 모델을 학습시키는 동작, 호흡 부전 발생 예측의 대상이 되는 제2 농약 중독 환자의 의료 데이터를 수신하는 동작, 상기 제2 농약 중독 환자의 의료 데이터에 대해 데이터 전처리를 수행하는 동작 및 상기 학습된 호흡 부전 예측 모델에 상기 데이터 전처리가 수행된 상기 제2 농약 중독 환자의 의료 데이터를 입력하는 것에 의해 상기 제2 농약 중독 환자의 호흡 부전 발생과 관련된 예측 데이터를 획득하는 동작을 포함할 수 있다.

[0011] 하드웨어와 결합되어 상기 농약 중독 환자의 호흡 부전 발생을 예측하는 예측 방법을 실행시키기 위하여 컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램.

## 도면의 간단한 설명

[0012] 도 1은 일 실시예에 따른 예측 장치가 농약 중독 환자의 호흡 부전 발생을 예측하는 과정을 설명하기 위한 도면이다.

도 2는 일 실시예에 따른 호흡 부전 발생을 예측하는 예측 장치의 구성들을 설명하기 위한 블록도이다.

도 3은 일 실시예에 따른 예측 장치가 결측 값을 보간하는 것을 설명하기 위한 도면이다.

도 4는 일 실시예에 따른 예측 장치가 시계열 데이터를 획득하는 것을 설명하기 위한 도면이다.

도 5는 일 실시예에 따른 호흡 부전 발생을 예측하는 예측 방법의 동작들을 설명하기 위한 흐름도이다.

### 발명을 실시하기 위한 구체적인 내용

- [0013] 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.
- [0014] 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.
- [0015] 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.
- [0016] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 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. 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.
- [0017] 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.
- [0018] 본 문서에서 사용된 용어 "모듈"은 하드웨어, 소프트웨어 또는 펌웨어로 구현된 유닛을 포함할 수 있으며, 예를 들면, 로직, 논리 블록, 부품, 또는 회로와 같은 용어와 상호 호환적으로 사용될 수 있다. 모듈은, 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는, 상기 부품의 최소 단위 또는 그 일부가 될 수 있다. 예를 들면, 일 실시예에 따르면, 모듈은 ASIC(application-specific integrated circuit)의 형태로 구현될 수 있다.
- [0019] 농약 중독 환자는 농약 중독 직후에는 아무런 증상을 나타내지 않고, 시간의 흐름에 따라 호흡 부전 등의 증상을 나타낼 수 있다. 농약 중독 환자를 치료하기 위해선 농약 중독 직후 아무런 증상이 나타나지 않을 때 정확한 초기 진단을 하는 것이 필요하다. 본원은 농약 중독 환자에 대한 초기 진단을 할 수 있도록 농약 중독 환자에 대한 호흡 부전을 예측할 수 있다. 본원은 농약 중독 환자가 중독 초기에는 호흡 부전이 없었으나 이후 호흡 부전이 발생할 것을 예측함으로써 농약 중독 환자가 적절한 치료를 받을 수 있도록 도울 수 있다. 도심권 병원의 의료진들이 농약 등의 약물 중독에 대한 진료 경험이 없거나 새로운 농약 약물에 관한 연구가 부족할 경우, 호흡 부전 발생 예측 모델을 통해 호흡 부전 발생을 예측하여 의료진이 조기 대응을 할 수 있도록 도울 수 있다.
- [0020] 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.
- [0022] 도 1은 일 실시예에 따른 예측 장치가 농약 중독 환자의 호흡 부전 발생을 예측하는 과정을 설명하기 위한 도면이다. 도1을 참조하면 예측 장치(예: 도2의 예측 장치(200))는 데이터를 수집, 학습 데이터를 정의, 호흡 부전 예측 모델을 학습시키고 대상 환자의 호흡 부전 발생을 예측할 수 있다.
- [0023] 일 실시예에서, 예측 장치는 환자의 의료 데이터를 수집(110)할 수 있다. 예를 들어 병원 공통 데이터 모델(common data model; CDM)로부터 환자의 의료 데이터를 수집할 수 있다. 의료 데이터는 농약 중독 환자의 기본 정보가 포함된 환자 정보 데이터, 생체 징후 정보를 포함하는 생체 징후 데이터 및 환자의 임상병리 검사 결과

가 포함된 랩 코드 데이터를 포함할 수 있다.

- [0024] 일 실시예에서, 예측 장치는 시계열 데이터에 기초하여 학습 데이터를 정의(120)할 수 있다. 시계열 데이터는 일정 시간 동안 수집되어 시간에 따른 정보를 포함하는 데이터를 나타낸다. 시계열 데이터는 예를 들어 수신된 의료 데이터에 포함된 항목들 중 최대 수축기 혈압, 최대 맥박, 최대 체온, 백혈구, 혈소판, 알부민, 총 CO<sub>2</sub>, C 반응 단백질, 산도, 산소 분압, 산소포화도, 젖산 수치를 포함할 수 있다. 일 실시예에서, 예측 장치는 시계열 데이터에 포함된 항목들 중 결측 값을 보간하고 슬라이딩 윈도우를 적용하여 복수의 시계열 데이터를 포함하는 학습 데이터를 결정할 수 있다.
- [0025] 일 실시예에서, 예측 장치는 시계열 데이터에 기초하여 정의된 학습 데이터 전부 또는 일부를 이용하여 호홉 부전 예측 모델을 학습시킬 수 있다. 학습 장치는 과대적합(overfitting) 문제를 방지하기 위하여 학습 데이터 중 일부는 호홉 부전 예측 모델을 학습시키고 나머지 일부로 호홉 부전 예측 모델을 테스트할 수 있다. 예를 들어, 예측 장치는 교차 검증 기법(예: stratified-k-fold)을 적용하여, 학습 데이터에 포함된 항목의 비율을 기초로 학습 데이터 중 일부만으로 호홉 부전 예측 모델을 학습시킬 수 있다. 일 실시예에서, 예측 장치는 학습 데이터를 이용하여 호홉 부전 예측 모델을 학습(130)시킬 수 있다. 예를 들어, 예측 장치는 학습 데이터의 수를 늘리고 학습 데이터의 불균형을 방지하기 위해 시계열 데이터에 ADASYN(adaptive synthetic; ADASYN) 오버샘플링 기법이 적용된 학습 데이터로 호홉 부전 예측 모델을 학습시킬 수 있다.
- [0026] 일 실시예에서, 예측 장치는 학습된 호홉 부전 예측 모델을 이용하여 대상 환자의 호홉 부전 발생을 예측(140)할 수 있다. 예를 들어, 예측 장치는 농약 중독 환자의 의료 데이터를 수신할 수 있다. 예측 장치는 농약 중독 환자의 의료 데이터에 대해 데이터 전처리를 수행하고, 전처리된 의료 데이터를 호홉 부전 예측 모델에 입력할 수 있다. 예측 장치는 호홉 부전 예측 모델로부터 농약 중독 환자의 호홉 부전과 관련된 예측 데이터를 획득하여 호홉 부전 발생을 예측할 수 있다. 일 실시예에서, 호홉 부전 예측 모델을 학습시키는 학습 과정은 모델의 학습을 위한 학습 장치에 의해 수행될 수도 있다. 이 경우, 예측 장치는 학습 장치에 의해 학습된 호홉 부전 예측 모델을 이용하여 대상 환자의 호홉 부전 발생을 예측할 수 있다.
- [0028] 도2는 일 실시예에 따른 호홉 부전 발생을 예측하는 예측 장치의 구성들을 설명하기 위한 블록 다이어그램이다. 도 2를 참조하면, 예측 장치(200)는 제어 모듈(210), 통신 모듈(220), 저장 모듈(230), 전처리 모듈(240), 학습 모듈(250) 및 예측 모듈(260)을 포함할 수 있다.
- [0029] 일 실시예에 따른 예측 장치(200)는 학습된 호홉 부전 예측 모델을 이용하여 농약 중독 환자의 호홉 부전 발생 여부를 예측하는 장치이다. 예측 장치(200)는 예측 장치에 포함된 제어 모듈(210), 통신 모듈(220), 저장 모듈(230), 전처리 모듈(240), 학습 모듈(250) 및 예측 모듈(260)을 이용하여 농약 중독 환자의 호홉 부전 발생 여부를 할 수 있다.
- [0030] 일 실시예에 따른 제어 모듈(210)은 적어도 하나의 프로세서를 구비할 수 있고, 예측 장치(200)에 포함된 하나 이상의 모듈을 제어할 수 있다. 예를 들어, 제어 모듈(210)은 통신 모듈(220)이 의료 데이터를 수신하고, 전처리 모듈(220)이 의료 데이터를 전처리하고, 학습 모듈(250)이 전처리된 의료 데이터를 호홉 부전 예측할 수 있도록 제어할 수 있다.
- [0031] 일 실시예에 따른 통신 모듈(220)은 농약 중독 환자의 의료 데이터를 수신할 수 있다. 통신 모듈(220)은 의료 데이터를 송신 또는 수신하기 위하여 송수신기를 포함할 수 있다. 예를 들어, 통신 모듈(220)은 농약 중독 환자의 환자 정보 데이터, 생체 징후 데이터 및 랩 코드 데이터를 포함하는 의료 데이터를 수신할 수 있다. 통신 모듈(220)은 사용자 단말 및 의료 기관의 서버 등으로부터 의료 데이터를 송수신할 수 있다. 통신 모듈은 지역 통신(Local Area Network; LAN) 모듈, 광역 통신(Wide Area Network; WAN)모듈, HDMI(High Definition Multimedia Interface), DVI(Digital Visual Interface), RS-232(recommended standard), 와이파이(Wifi) 모듈, 와이브로(Wireless broadband) 모듈, GSM(global System for Mobile Communication), CDMA(Code Division Multiple Access), WCDMA(Wideband Code Division Multiple Access), UMTS(universal mobile telecommunications system), TDMA(Time Division Multiple Access), WLAN(Wireless LAN), DLNA(Digital Living Network Alliance), WiBro(Wireless Broadband), WiMAX(World Interoperability for Microwave Access), LTE(Long Term Evolution), 5G등을 포함할 수 있다.
- [0032] 일 실시예에 따른 저장 모듈(230)은 예측 장치(200)가 농약 중독 환자의 호홉 부전 여부를 예측하기 위해 수신된 의료 데이터, 어플리케이션 및 프로세서에 의해 수행될 수 있는 명령어들을 저장할 수 있다. 예를 들어, 저

장 모듈(230)은 농약 호흡 부전 예측 모델을 메모리에 저장할 수 있고, 예측 모델은 로지스틱 회귀 모델, 결정 트리 모델, 랜덤 포레스트 모델 및 LSTM(Long Short-Term Memory) 모델을 포함할 수 있다. 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.

[0033] 일 실시예에 따른 전처리 모듈(240)은 수신된 의료 데이터를 전처리할 수 있다. 전처리 모듈(240)은 문자열로 저장된 값을 숫자로 변환하는 원 핫 인코딩 방식, 데이터 스케일링, 데이터 오버 샘플링, 시계열 데이터에 대한 전처리 등을 수행할 수 있다. 시계열 데이터에 대한 전처리는 의료 데이터의 결측 값에 대한 보간 및 슬라이딩 윈도우에 기반하여 복수의 시계열 데이터를 획득하는 것을 포함할 수 있다. 의료 데이터에 대한 결측 값을 보간하는 것은 도 3에서 상세히 설명하고 슬라이딩 윈도우에 기반하여 복수의 시계열 데이터를 획득하는 것은 도 4에서 상세히 설명하도록 한다.

[0034] 일 실시예에 따른 예측 장치는 의료 데이터에 대하여 데이터 스케일링을 할 수 있다. 예측 장치는 획득된 시계열 데이터에 포함된 각 항목에 대한 최대 값 및 최소 값을 기초로 각 항목의 값들을 스케일링할 수 있다. 예를 들어 예측 장치는 최대-최소 스케일링(min-max scaling) 기법을 이용하여 데이터를 스케일링 할 수 있다. 최대-최소 스케일링 기법을 나타내는 식은 하기 수학적 식 1과 같다.

### 수학적 식 1

$$X_s = \frac{X - \text{Min}}{\text{Max} - \text{Min}}$$

[0035]

[0036] 여기서, X는 스케일링 되기 전 데이터, Xs는 스케일링 된 후의 데이터, Min은 항목에 포함된 데이터의 최소 값, Max는 항목에 포함된 데이터의 최대 값을 나타낸다. 의료 데이터에 포함된 데이터의 값의 범위는 항목마다 다르기 때문에 호흡 부전 예측 모델은 큰 값이 포함되어 있는 데이터의 항목에 편향될 수 있다. 최대-최소 스케일링 기법을 통해 스케일링 된 데이터에 포함된 값의 범위는 0과 1사이의 값으로 변환될 수 있다.

[0037] 오버 샘플링 기법을 제외한 데이터 전처리 후의 데이터는 호흡 부전 예측 모델을 학습시키기 위한 학습 데이터와 호흡 부전 예측 모델을 평가하기 위한 테스트 데이터로 나누어질 수 있다. 학습 데이터는 시계열 특성이 없는 항목들(예: 음독량, 성별, 나이, 및 의식상태와 시계열 특성이 있는 항목들(예: 1시간 동안의 최대 수축기 혈압, 최대 맥박, 최대 체온 및 3시간 간격으로 측정된 백혈구, 혈소판, 알부민, 총 CO2(total CO2; TC02), C반응 단백질, 산도, 산소 분압, 산소포화도, 젖산의 농도)을 포함할 수 있다.

[0038] 농약 중독 환자의 데이터에서 호흡 부전 환자의 비율에 따라 데이터 불균형 문제가 발생할 수 있다. 오버샘플링 기법은 농약 중독 환자 데이터의 양이 많지 않고 실험적으로 가장 좋은 성능을 나타내는 오버샘플링 기법이 사용될 수 있다. 일 실시예에 따른 예측 장치는 시계열 데이터에 오버샘플링 기법을 적용하여 학습 데이터를 생성할 수 있다. 예를 들어, 예측 장치는 획득된 복수의 시계열 데이터에 ADASYN 오버샘플링 기법을 적용하여 학습 데이터 세트를 생성할 수 있다. ADASYN 오버샘플링 기법은 학습하기 어려운 데이터를 학습하기 쉽도록 데이터를 생성하는 기법이다. ADASYN 오버샘플링 기법은 소수 항목에 포함된 데이터에 k-최근접 이웃 알고리즘을 적용하여 소수 항목에 포함된 데이터의 근처에 다수 항목에 포함된 데이터가 많을수록 소수 항목에 포함된 데이터를 많이 생성할 수 있다.

[0039] 일 실시예에서, ADASYN 오버샘플링 기법은 4가지 단계를 포함할 수 있다. 첫 번째, 모든 소수 항목에 포함된 데이터는 k-최근접 이웃 알고리즘을 적용하여 다수 항목에 포함된 데이터에 대한 제1 값을 구한다. 두 번째, 첫 번째 단계에서 구한 제1 값의 총합인 제2 값을 구한다. 세 번째, 제2 값에 (다수 항목에 포함된 데이터 개수 - 소수 항목에 포함된 데이터 개수)에 해당하는 수를 곱하고, 소수점 자리는 버린 제3 값을 구한다. 제3 값은 소수 항목에 포함된 새로운 데이터가 생성되는 양이다. 마지막으로, 제3 값을 기초로 모든 항목에 대해 소수 합성 오버샘플링 기법(synthetic minority over-sampling technique; SMOTE)을 적용할 수 있다. 학습 장치는 소수 항목에 포함된 데이터 및 k-최근접 이웃 알고리즘에 기초한 소수 합성 오버샘플링 기법이나 다수 항목과 소수

항목 사이의 경계선에 소수 합성 오버샘플링 기법을 적용하는 경계선- 소수 합성 오버샘플링 기법(Borderline-SMOTE)을 적용하여 시계열 데이터를 오버샘플링 할 수도 있다. 전처리 된 데이터는 2차원 데이터로 변환되어 호홉 부전 예측 모델의 학습에 사용될 수 있다.

[0040] 일 실시예에 따른 학습 모듈(250)은 획득된 복수의 시계열 데이터를 기초로 농약 중독 환자의 호홉 부전 발생을 예측하는 호홉 부전 예측 모델을 학습시킬 수 있다. 학습 모듈이 학습시키는 호홉 부전 예측 모델은 로지스틱 회귀 모델, 결정 트리 모델, 랜덤 포레스트 모델 및 LSTM(long short-term memory; LSTM) 모델 중 적어도 하나를 포함하는 것일 수 있다.

[0041] 일 실시예에 따른, 로지스틱 회귀 모델(logistic regression; LR)은 기계 학습 분류 모델 중 하나이다. 로지스틱 회귀 모델은 시그모이드(sigmoid) 함수를 이용하여 이진 분류를 진행할 수 있다. 시그모이드 함수는 하기 수학적 식 2로 나타낼 수 있다.

### 수학적 식 2

$$f(z) = \frac{1}{1 + e^{-z}}$$

[0042]

[0043] 시그모이드 함수에 들어갈 인자는 학습 데이터를 기반으로 가중치와 바이어스를 조절하여 계산될 수 있다. 예를 들어, 항목의 개수가 41개인 경우 하기 수학적 식 3로 나타낼 수 있다.

### 수학적 식 3

$$z = \sum_{i=1}^{41} w_i a_i + b$$

[0044]

[0045] 수학적 식 3을 통해 만들어진 z값은 수학적 식 2의 시그모이드 함수에 입력될 수 있다.

[0046] 이진 분류는 하기 수학적 식 4에 의하여 수행될 수 있다. 시그모이드 함수  $f(z)$ 의 값이 0.5보다 작거나 같으면 음성, 0.5보다 크면 양성으로 구분될 수 있다.

### 수학적 식 4

$$predict = \begin{cases} 0 & \text{if } f(z) \leq 0.5 \\ 1 & \text{if } f(z) > 0.5 \end{cases}$$

[0047]

[0049] 결정 트리(decision tree; DT)는 호홉 부전 발생 여부와 관련 있는 항목에 대한 값을 통해 분기하는 방식으로 분류를 진행할 수 있다. 결정 트리는 은 노드(node)의 샘플(sample)에서 불순도를 측정하여 분기하고, 불순도 값이 작아지도록 한다. 불순도를 측정하는 방식은 지니(Gini) 지수와 엔트로피 지수를 사용할 수 있다. 지니 지수를 사용하여 불순도를 측정하는 경우, 하기 지니 지수는 수학적 식 5로 나타낼 수 있다.

수학식 5

$$\text{Gini} = \sum_{i=1}^n (R_i(1 - \sum_{k=1}^m P_k^2))$$

[0050]

[0051]

$R_i$ 는 분기된 데이터 개수의 비율,  $P_k$ 는 항목에 해당하는 데이터 비율이다. 결정 트리가 데이터에 과도하게 의존하는 것을 막기 위해 가지치기(Pruning)가 수행될 수 있다. 예를 들어, 결정 트리 분류기(decision tree classifier)의 최대 깊이를 6으로 가지치기가 수행될 수 있다.

[0052]

일 실시예에 따른 랜덤 포레스트(random forest; RF) 모델은 앙상블 기법의 배깅(bagging) 과정을 통해 결정 트리가 무작위 샘플링 데이터를 기초로 학습되고, 학습 데이터를 학습된 모든 결정 트리의 조합으로 분류하는 모델이다. 예를 들어, 랜덤 포레스트에서 만들어진 결정 트리에 분류할 데이터가 통과되고, 각 결정 트리 결과는 해당 클래스에 투표하여 투표를 가장 많이 받은 클래스가 최종 선택될 수 있다. 결정 트리는 100개이고 최대 깊이는 6일 수 있다.

[0053]

일 실시예에 따른 롱 쇼트-텀 메모리(long short-term memory; LSTM) 모델은 입력 데이터의 장기 의존성 문제를 해결하기 위한 순환 신경망(recurrent neural network; RNN)의 유형 중 하나이다. 순환 신경망은 자신의 출력을 다시 입력으로 사용할 수 있다. 롱 쇼트-텀 메모리가 적용된 순환 신경망 모델은 장기 의존성 문제를 해결하여 긴 시퀀스를 포함하는 시계열 데이터를 학습시킬 수 있다.

[0054]

일 실시예에 따른 예측 모듈(260)은 학습된 호흡 부전 예측 모델에 데이터 전처리가 수행된 농약 중독 환자의 의료 데이터를 입력하고, 호흡 부전 예측 모델로부터 농약 중독 환자의 호흡 부전 발생과 관련된 예측 데이터를 획득할 수 있다. 예를 들어, 예측 모듈(260)은 호흡 부전 예측 모델로부터 농약 중독 환자의 호흡 부전 발생과 관련된 민감도(sensitivity), 양성 예측도(positive predictive value; PPV) 및 정확도를 포함한 예측 데이터를 획득할 수 있다. 양성 예측도는 호흡 부전 예측 모델이 호흡 부전 환자로 예측한 호흡 부전 환자 중 실제 호흡 부전 환자의 비율이고, 민감도는 실제 호흡 부전 환자 중 호흡 부전 예측 모델이 호흡 부전 환자로 예측한 비율이다.

[0055]

도 3은 일 실시예에 따른 예측 장치가 결측 값을 보간하는 것을 설명하기 위한 도면이다.

[0056]

도3을 참조하면, 예측 장치(예: 도2의 예측 장치(200))는 최근 데이터 보간 방법으로 결측 값을 보간할 수 있다. 예를 들어, 예측 장치는 수신된 의료 데이터의 시계열 특성을 기초로 결측 값이 포함된 항목의 가장 최근에 측정된 값, 수신한 의료 데이터에서 결측 값이 포함된 항목과 연관성이 가장 높은 항목의 측정 값 및 결측 값이 포함된 항목의 측정 값들의 평균 값 중 어느 하나를 기초로 결측 값을 보간할 수 있다. 최근 데이터 보간 기법(recent data imputation; RDI)은 시계열 데이터의 특성에 기초하여 결측 값을 최신 데이터로 보간하는 방법이다.

[0057]

일 실시예에 따른 시계열 데이터는 1 시간 간격으로 측정된 의료 데이터에 포함된 시계열 데이터일 수 있다. 예를 들어, 시계열 데이터(310)은 1시간 간격으로 측정된 시계열 데이터 중 2시간이 되는 시점에 측정된 시계열 데이터일 수 있다. 일 실시예에서, 예측 장치는 결측 값이 포함된 항목의 가장 최근에 측정된 값으로 결측 값을 보간할 수 있다. 예를 들어 예측 장치는 시계열 데이터(310)에 포함된 결측 값(320)을 3 시간이 되는 시점에 측정된 측정 값(330)으로 보간할 수 있다. 예측 장치는 시계열 데이터(310)에 포함된 측정 값(340)으로 1시간이 되는 시점에 측정된 시계열 데이터에 포함된 결측 값(350)을 보간할 수 있다.

[0058]

일 실시예에 따른 예측 장치는 수신한 의료 데이터에서 결측 값이 포함된 항목과 연관성이 가장 높은 항목의 측정 값을 기초로 결측 값을 보간할 수 있다. 예를 들어, 예측 장치는 연관성이 높은 항목들의 거리 함수를 기초로 결측 값을 보간하는 K-최근접 이웃 보간(k-nearest neighbor imputer; KNN Imputer) 기법을 적용하여 결측 값을 보간할 수 있다. 농약 중독 환자의 의료 데이터에 포함된 항목들의 제1 집합은 총 이산화 탄소(total CO<sub>2</sub>; TC0<sub>2</sub>), 산도(pH), 중탄산염(HCO<sub>3</sub>std), 베이스 엑세스(base excess; BE), 젯산에 관한 것이고, 제2 집합은 이산화탄소분압(pCO<sub>2</sub>), 산소분압(pO<sub>2</sub>)에 관한 것일 수 있다. 예측 장치는 각각의 집합에 대하여 K-최근접 이웃 보간 기법을 적용할 수 있다. 일 실시예에 따른 예측 장치는 최근 데이터 보간 기법 및 K-최근접 이웃 보간 기법을

적용 후 남은 결측 값에 대하여 측정 값 및 결측 값이 포함된 항목의 측정 값들의 평균 값으로 결측 값을 보간할 수 있다.

- [0060] 도 4는 일 실시예에 따른 예측 장치가 시계열 데이터를 획득하는 것을 설명하기 위한 도면이다.
- [0061] 도 4를 참조하면, 예측 장치(예: 도2의 예측 장치(200))는 농약 중독 환자들에 대한 복수의 시계열 데이터를 획득할 수 있다. 예를 들어, 예측 장치는 환자(410, 420, 430)으로부터 복수의 시계열 데이터(413, 423, 433)를 획득할 수 있다.
- [0062] 일 실시예에 따른, 예측 장치는 농약 중독 환자의 하루 단위의 생체 징후 데이터에 미리 정의된 시간 단위의 슬라이딩 윈도우(sliding window)를 시간의 흐름에 따라 순차적으로 적용하여 농약 중독 환자에 대한 복수의 시계열 데이터를 획득할 수 있다. 예를 들어, 미리 정의된 시간 단위가 3 시간인 슬라이딩 윈도우를 순차적으로 적용하는 경우, 농약 중독 환자로부터 최소 0개에서 최대 22개의 시계열 데이터를 획득할 수 있다. 예측 장치는 호흡 부전이 발생한 환자(410)의 시계열 데이터에 슬라이딩 윈도우(411)부터 슬라이딩 윈도우(412)까지 순차적으로 적용하는 경우 22개의 시계열 데이터(413)를 획득하고, 환자(430)의 시계열 데이터에 슬라이딩 윈도우(431)부터 슬라이딩 윈도우(432)까지 순차적으로 적용하는 경우 10개의 시계열 데이터(433)를 획득할 수 있다. 예측 장치는 호흡 부전이 발생하지 않은 환자(420)의 시계열 데이터에 슬라이딩 윈도우(421)부터 슬라이딩 윈도우(422)까지 순차적으로 적용하는 경우 22개의 시계열 데이터(423)를 획득할 수 있다.
- [0063] 일 실시예에 따른, 획득된 복수의 시계열 데이터에 포함된 항목들은 음독 농약, 음독량, 성별, 나이, 의식 상태, 혈압, 맥박, 체온, 백혈구 수치, 혈소판 수치, 알부민 수치, 총 이산화탄소 수치, C-반응 단백질 수치, 산도, 산소 분압, 산소 포화도 및 젖산 수치를 포함할 수 있다. 시계열 데이터에 포함될 항목을 선별함으로써 의료 데이터의 전처리 과정에서 호흡 부전 예측 모델의 예측 정확도를 높일 수 있다.
- [0064] 일 실시예에 따른 예측 장치는 호흡부전과 관련 있는 항목 및 항목에 대한 유의 확률(p-value)을 추출할 수 있다. 예측 장치는 유의 확률이 0.05보다 큰 항목을 추출하여 시계열 데이터에 포함하도록 결정할 수 있다. 예를 들어, 예측 장치는 전체 항목들 중 유의 확률이 0.05보다 큰 음독 농약, 음독량, 성별, 나이, 의식 상태, 혈압, 맥박, 체온, 백혈구 수치, 혈소판 수치, 알부민 수치, 총 이산화탄소 수치, C-반응 단백질수치, 산도, 산소 분압, 산소 포화도 및 젖산 수치의 전부 또는 일부를 포함할 수 있다.
- [0065] 일 실시예에 따른 예측 장치는 과대 적합(overfitting)을 방지하기 위해 복수의 시계열 데이터 중 일부를 학습 데이터 세트로, 나머지 일부를 훈련 데이터 세트로 나눌 수 있다. 예를 들어, 예측 장치는 k 폴드 교차 검증(예: stratified-k-fold) 기법을 시계열 데이터에 적용하여 각 항목의 비율에 기초한 비율(예: 7:3)로 학습 데이터 세트와 훈련 데이터 세트를 나눌 수 있다.
- [0067] 도 5는 일 실시예에 따른 호흡 부전 발생을 예측하는 예측 방법의 동작들을 설명하기 위한 흐름도이다. 호흡 부전 발생을 예측하는 예측 방법의 동작들은 본 개시에서 설명된 예측 장치(예: 도 2의 예측 장치(200))에 의해 수행될 수 있다.
- [0068] 동작(510)에서, 예측 장치는 제1 농약 중독 환자의 의료 데이터를 수신할 수 있다. 예를 들어, 예측 장치는 병원 공통 데이터 모델로부터 호흡 부전 예측 모델을 학습시키기 위한 제1 농약 중독 환자의 의료 데이터를 수신할 수 있다. 제1 농약 중독 환자의 의료 데이터는 환자의 환자 정보 데이터, 환자의 생체 징후 정보를 포함하는 생체 징후 데이터 및 임상병리 검사 결과 정보를 포함하는 랩 코드 데이터를 포함할 수 있다.
- [0069] 동작(520)에서, 예측 장치는 결측 값이 보간된 제1 농약 중독 환자의 의료 데이터를 생성할 수 있다. 예측 장치는 수신한 의료 데이터의 결측 값을 보간할 수 있다. 일 실시예에서 예측 장치는 수신된 의료 데이터의 시계열 특성을 기초로 결측 값이 포함된 항목의 가장 최근에 측정된 값, 수신한 의료 데이터에서 결측 값이 포함된 항목과 연관성이 가장 높은 항목의 측정 값 및 결측 값이 포함된 항목의 측정 값들의 평균 값 중 어느 하나를 기초로 결측 값을 보간할 수 있다. 예를 들어 예측 장치는 최근 데이터 보간 기법을 적용하여 결측 값을 보간하고, 남은 결측 값에 대하여 k-최근접 이웃 보간 기법을 적용한 뒤 결측 값이 포함된 항목의 평균 값으로 남은 결측 값을 보간할 수 있다.
- [0070] 동작(530)에서, 예측 장치는 결측 값이 보간된 제1 농약 중독 환자의 의료 데이터로부터 복수의 시계열 데이터를 획득할 수 있다. 일 실시예에서 예측 장치는 미리 정의된 시간 단위의 슬라이딩 윈도우를 시간의 흐름에 따

라 순차적으로 적용하여 제1 농약 중독 환자에 대한 복수의 시계열 데이터를 획득할 수 있다. 예를 들어, 예측 장치는 미리 정의된 시간 단위가 3 시간인 슬라이딩 윈도우를 순차적으로 적용하여 농약 중독 환자로부터 최소 0개에서 최대 22개의 시계열 데이터를 획득할 수 있다.

[0071] 일 실시예에 따른 예측 장치는 각 항목에 대한 유의 확률을 기초로 획득된 복수의 시계열 데이터에 포함된 항목들을 결정할 수 있다. 예를 들어, 예측 장치는 유의 확률이 0.05보다 큰 항목인 음독 농약, 음독량, 성별, 나이, 의식 상태, 혈압, 맥박, 체온, 백혈구 수치, 혈소판 수치, 알부민 수치, 총 이산화탄소 수치, C-반응 단백질 수치, 산도, 산소 분압, 산소 포화도 및 젖산 수치를 시계열 데이터에 포함되도록 결정할 수 있다.

[0072] 일 실시예에서, 예측 장치는 획득된 시계열 데이터에 포함된 각 항목에 대한 최대 값 및 최소 값을 기초로 각 항목의 값들을 스케일링할 수 있다. 예를 들어 예측 장치는 최대-최소 스케일링(min-max scaling) 기법을 적용하여 획득된 시계열 데이터를 스케일링 할 수 있다.

[0073] 동작(540)에서, 예측 장치는 제1 농약 중독 환자에 관한 복수의 시계열 데이터를 이용하여 호흡 부전 예측 모델을 학습시킬 수 있다. 예측 장치는 호흡 부전 예측 모델의 호흡 부전 예측 정확도를 향상시키기 위하여 ADASYN 오버샘플링 기법, 경계선-소수 합성 오버샘플링 기법, 소수 합성 오버샘플링 기법 중 어느 한 기법을 이용하여 시계열 데이터를 오버 샘플링 할 수 있다. 예를 들어, 예측 장치는 획득된 복수의 시계열 데이터에 ADASYN 오버샘플링 기법을 적용하여 생성된 학습 데이터 세트를 상기 호흡 부전 예측 모델에 학습시킬 수 있다.

[0074] 동작(550)에서, 예측 장치는 제2 농약 중독 환자의 의료 데이터를 수신할 수 있다. 제2 농약 중독 환자의 의료 데이터는 호흡 부전 발생 예측의 대상이 되는 농약 중독 환자의 의료 데이터이다. 예를 들어, 예측 장치는 농약 중독 환자에 대한 치료 경험이 적은 도심권 병원으로부터 호흡 부전 발생 예측의 대상이 되는 제2 농약 중독 환자의 의료 데이터를 수신할 수 있다.

[0075] 동작(560)에서, 예측 장치는 제2 농약 중독 환자의 의료 데이터를 전처리할 수 있다. 제2 농약 중독 환자의 의료 데이터는 환자 정보 데이터, 생체 징후 데이터 및 랩 코드 데이터를 포함할 수 있다. 예측 장치는 제2 농약 중독 환자의 호흡 부전을 예측하기 위하여 의료 데이터를 전처리 할 수 있다. 예를 들어 예측 장치는 의료 데이터에 결측 값 보간, 윈도우 슬라이딩을 적용한 시계열 데이터의 생성, 시계열 데이터에 대한 스케일링 등을 포함한 전처리를 수행할 수 있다.

[0076] 동작(570)에서, 예측 장치는 학습된 호흡 부전 예측 모델을 이용하여 제2 농약 중독 환자의 호흡 부전 발생과 관련된 예측 데이터를 획득할 수 있다. 예측 장치는 학습된 호흡 부전 예측 모델에 데이터 전처리가 수행된 제2 농약 중독 환자의 의료 데이터를 입력하고, 제2 농약 중독 환자의 호흡 부전 발생과 관련된 예측 데이터를 획득할 수 있다. 예를 들어 예측 장치는 제2 농약 중독 환자의 전처리된 데이터를 로지스틱 회귀모델, 결정 트리 모델, 랜덤 포레스트 모델 및 LSTM 모델 중 적어도 하나를 포함하는 호흡 부전 예측 모델에 입력하고, 호흡 부전 발생과 관련된 예측 데이터를 획득 할 수 있다. 호흡 부전과 관련된 예측 데이터는 호흡 부전 발생과 관련된 민감도(sensitivity), 양성 예측도(positive predictive value; PPV) 및 정확도를 포함할 수 있다.

[0078] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0079] 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 처

리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 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.

[0080] 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 저장할 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.

[0081] 위에서 설명한 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 또는 복수의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

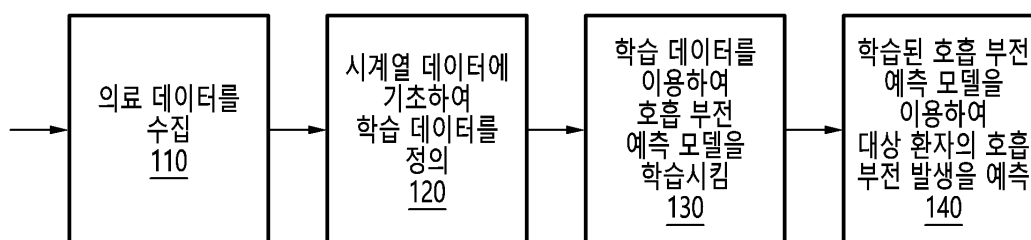
[0082] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 이를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[0083] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

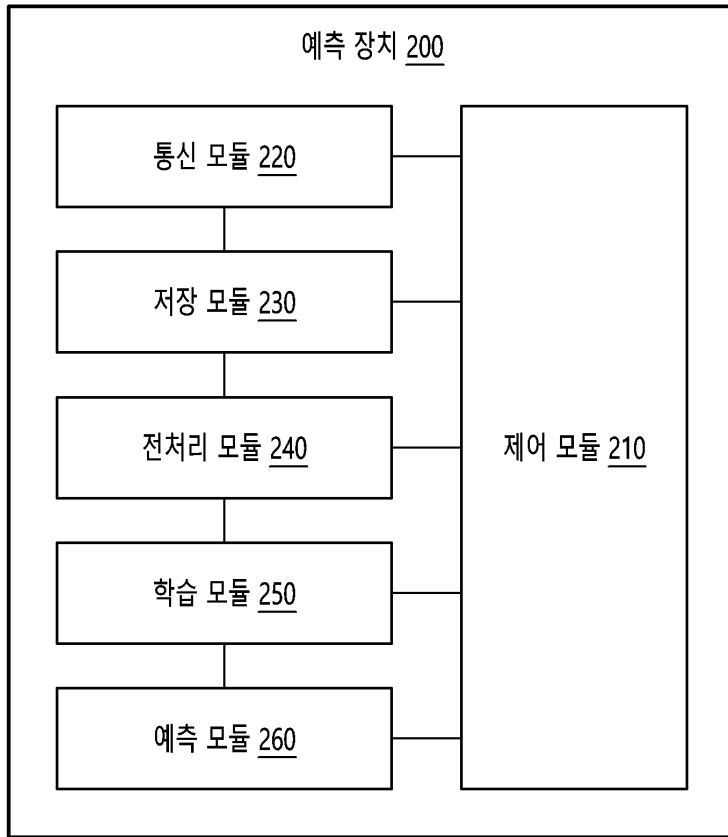
## 부호의 설명

### 도면

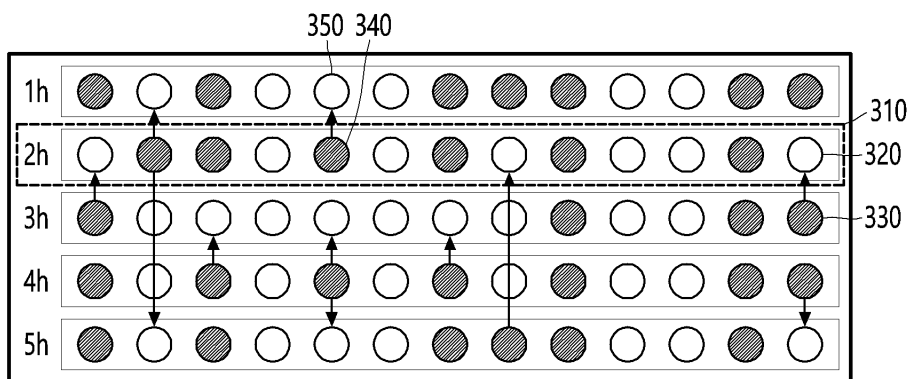
#### 도면1



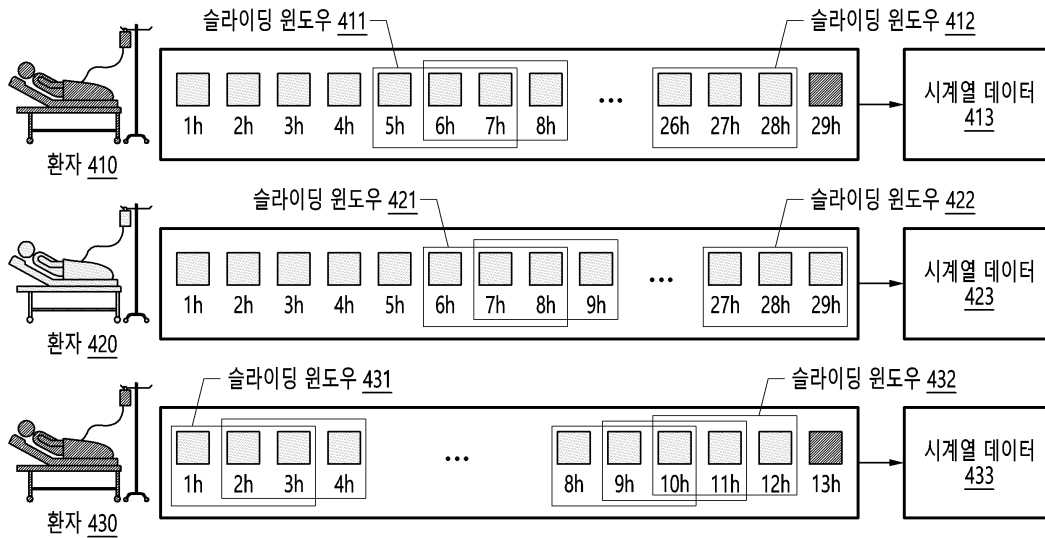
도면2



도면3



도면4



도면5

