



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0038629  
(43) 공개일자 2025년03월19일

(51) 국제특허분류(Int. Cl.)

G16H 50/70 (2018.01) A61B 5/00 (2021.01)  
G06N 3/045 (2023.01) G16B 20/20 (2019.01)  
G16B 30/10 (2019.01) G16H 10/20 (2018.01)  
G16H 10/60 (2018.01) G16H 20/10 (2018.01)  
G16H 50/20 (2018.01) G16H 50/50 (2018.01)

(52) CPC특허분류

G16H 50/70 (2018.01)  
A61B 5/4848 (2013.01)

(21) 출원번호 10-2024-0124584

(22) 출원일자 2024년09월12일

심사청구일자 2024년09월12일

(30) 우선권주장

1020230120867 2023년09월12일 대한민국(KR)

(71) 출원인

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

주식회사 온코마스터

서울특별시 동대문구 회기로 117-3, 산업지원동 305호(회기동)

(72) 발명자

김좌훈

서울특별시 구로구 신도림로 16, 501동 803호 (신도림동, 신도림 대림아파트)

박경화

서울특별시 성북구 성북로 176, 101동 102호 (성북동, 성북동외교관사택단지)

(뒷면에 계속)

(74) 대리인

특허법인태백

전체 청구항 수 : 총 12 항

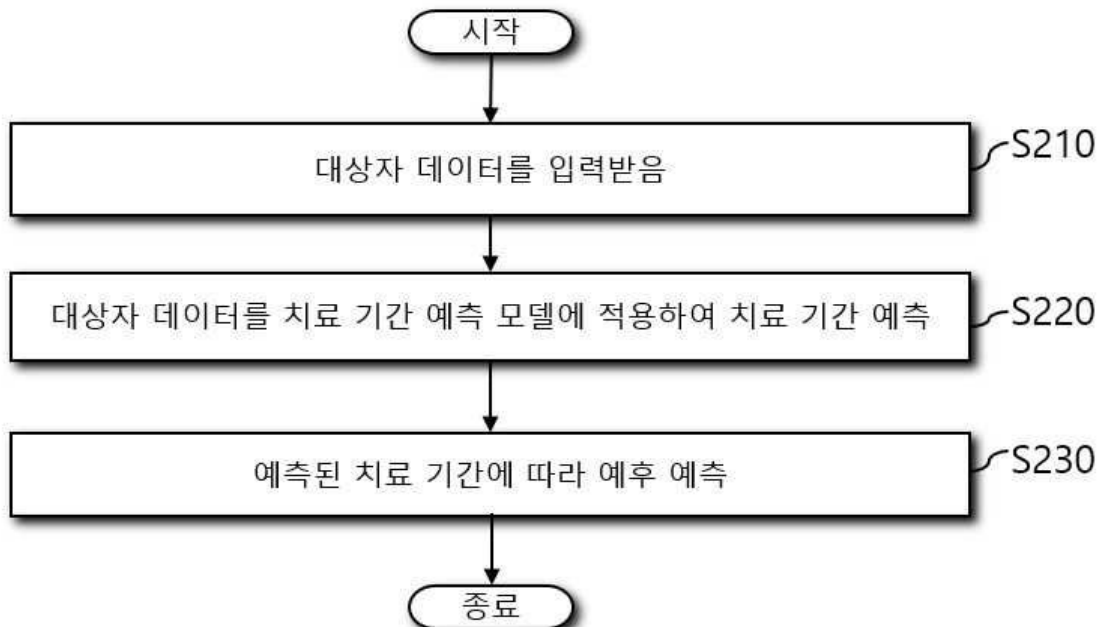
(54) 발명의 명칭 위암 환자의 2차 치료제 예후 예측 방법 및 그 시스템

(57) 요약

본 발명은 위암 환자의 2차 치료제 예후 예측 방법 및 그 시스템에 관한 것이다. 본 발명에 따르면 2차 치료제 예후 예측 방법에 있어서, 대상자 데이터 입력부가 대상자의 임상 정보 및 유전자 정보를 포함하는 대상자 데이터를 입력받는 단계; 치료 기간 예측부가 입력된 상기 대상자 데이터를 기 구축된 치료 기간 예측 모델에 적용하

(뒷면에 계속)

대표도 - 도2



여 기 설정된 하나 이상의 2차 치료제 각각에 대한 치료 기간을 예측하는 단계; 및 예후 예측부가 예측된 상기 치료 기간에 따라 대상자의 2차 치료제에 대한 반응 및 예후를 예측하는 단계를 포함할 수 있다.

이와 같이 본 발명에 따르면, 2차 치료제에 대한 환자의 치료 기간을 예측하여 체계적으로 환자의 치료 계획을 수립하는데 활용될 수 있다.

또한, 종래의 항암화학요법에 대한 반응성을 예측하기 위한 방법들은 모든 약물의 수의 제공에 대하여 실험이 요구되므로 시간, 비용 및 노동력 측면에 있어서 매우 비효율적이었으나 대상자의 임상 정보 및 유전자 정보를 기 구축된 인공지능 모델에 적용하여 치료 기간을 예측함으로써 신속하고 간편하게 2차 치료제 치료에 대한 반응 및 예후를 신뢰성있게 예측할 수 있다.

(52) CPC특허분류

*G06N 3/045* (2023.01)  
*G16B 20/20* (2019.02)  
*G16B 30/10* (2019.02)  
*G16H 10/20* (2021.08)  
*G16H 10/60* (2021.08)  
*G16H 20/10* (2021.08)  
*G16H 50/20* (2018.01)  
*G16H 50/50* (2018.01)

**최용화**

경기도 수원시 영통구 창룡대로 250, 101동 205호  
 (이의동, 수원광교 행복주택)

(72) 발명자

**최윤지**

서울특별시 강동구 상일로 74, 339동 1502호 (상일  
 동, 고덕리엔파크3단지아파트)

이 발명을 지원한 국가연구개발사업

|             |                             |
|-------------|-----------------------------|
| 과제고유번호      | 1465039933                  |
| 과제번호        | HR22C1302050023             |
| 부처명         | 보건복지부                       |
| 과제관리(전문)기관명 | 한국보건산업진흥원                   |
| 연구사업명       | 연구중심병원 육성 R&D               |
| 연구과제명       | 디지털 트윈 기반 정밀의료 통합 관리 플랫폼 구현 |
| 과제수행기관명     | 고려대학교 안암병원                  |
| 연구기간        | 2023.01.01 ~ 2023.12.31     |

## 명세서

### 청구범위

#### 청구항 1

대상자 데이터 입력부가 대상자의 임상 정보 및 유전자 정보를 포함하는 대상자 데이터를 입력받는 단계;

치료 기간 예측부가 입력된 상기 대상자 데이터를 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료제 각각에 대한 치료 기간을 예측하는 단계; 및

예후 예측부가 예측된 상기 치료 기간에 따라 대상자의 2차 치료제에 대한 반응 및 예후를 예측하는 단계를 포함하는 2차 치료제 예후 예측 방법.

#### 청구항 2

제1항에 있어서,

상기 임상 정보는,

대상자의 성별, 나이, 조직학적 특성, 전이 및 재발 여부, 1차 치료에서 표적치료제 사용 여부 및 1차 치료제 치료 기간을 포함하고,

상기 유전자 정보는,

대상자의 신체 일 부위의 암에 대한 병원성 유전자의 단일 염기서열 변이 및 유전자 복제 변이 유무와 발암성 유전자의 단일 염기서열 변이 및 유전자 복제 변이 유무에 대한 정보를 포함하는 2차 치료제 예후 예측 방법.

#### 청구항 3

제1항에 있어서,

상기 치료 기간을 예측하는 단계는,

입력된 대상자 데이터 중 유전자 정보에 기 설정된 임베딩 방법을 통해 벡터 형태로 인코딩하고 상기 임상 정보와 임베딩된 상기 유전자 정보를 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료제 각각에 대한 치료 기간을 병렬적으로 예측하는 2차 치료제 예후 예측 방법.

#### 청구항 4

제3항에 있어서,

상기 치료 기간을 예측하는 단계는,

상기 임상 정보 및 유전자 정보를 각각 0 또는 1로 이진화하고,

모든 유전자에 대한 임베딩 벡터가 특업 테이블 형태로 저장된 임베딩 벡터 정보 테이블에서 상기 유전자 정보 중 1에 해당하는 유전자 정보의 임베딩 벡터를 추출하여 합산하며,

합산된 상기 유전자 정보를 이진화된 상기 임상 정보와 결합하여 기 구축된 치료 기간 예측 모델에 적용하여 2차 치료제에 대한 치료 기간을 각각 예측하는 2차 치료제 예후 예측 방법.

#### 청구항 5

제1항에 있어서,

상기 반응 및 예후를 예측하는 단계는,

예측된 치료 기간이 제1 기간 미만인 경우 2차 치료제에 대한 반응 없음으로 예측하고,

예측된 치료 기간이 제1 기간 이상이고 제2 기간 미만이면 2차 치료제에 대한 반응 알 수 없음으로 예측하며,

예측된 치료 기간이 제2 기간 이상이면 2차 치료제에 대한 반응 있음으로 예측하는 2차 치료제 예후 예측 방법.

#### 청구항 6

제2항에 있어서,

상기 치료 기간을 예측하는 단계는,

유전자에서 EZH2, FGFR1, CALR, TOP1, SRSF2, PIK3CA, ALK 중 적어도 하나에 유전적 변이가 있거나 조직학적 특성이 미만형인 경우 2차 치료제의 치료 기간을 제2 기간 이상으로 예측하고,

유전자에서 EGFR, SMARCA4, MTOR, APC, PBRM1, FBXW7, EGFR 중 적어도 하나에 유전적 변이가 있거나 초기에 암이 전이된 환자의 경우 2차 치료제의 치료 기간이 제1 기간 미만으로 예측하는 2차 치료제 예후 예측 방법.

#### 청구항 7

대상자의 임상 정보 및 유전자 정보를 포함하는 대상자 데이터를 입력받는 대상자 데이터 입력부;

입력된 상기 대상자 데이터를 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료제 각각에 대한 치료 기간을 예측하는 치료 기간 예측부; 및

예측된 상기 치료 기간에 따라 대상자의 2차 치료제에 대한 반응 및 예후를 예측하는 예후 예측부를 포함하는 2차 치료제 예후 예측 시스템.

#### 청구항 8

제7항에 있어서,

상기 임상 정보는,

대상자의 성별, 나이, 조직학적 특성, 전이 및 재발 여부, 1차 치료에서 표적치료제 사용 여부 및 1차 치료제 치료 기간을 포함하고,

상기 유전자 정보는,

대상자의 신체 일 부위의 암에 대한 병원성 유전자의 단일 염기서열 변이 및 유전자 복제 변이 유무와 발암성 유전자의 단일 염기서열 변이 및 유전자 복제 변이 유무에 대한 정보를 포함하는 2차 치료제 예후 예측 시스템.

#### 청구항 9

제7항에 있어서,

상기 치료 기간 예측부는,

입력된 대상자 데이터 중 유전자 정보에 기 설정된 임베딩 방법을 통해 벡터 형태로 인코딩하고 상기 임상 정보와 임베딩된 상기 유전자 정보를 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료제 각각에 대한 치료 기간을 병렬적으로 예측하는 2차 치료제 예후 예측 시스템.

#### 청구항 10

제9항에 있어서,

상기 치료 기간 예측부는,

상기 임상 정보 및 유전자 정보를 각각 0 또는 1로 이진화하고,

모든 유전자에 대한 임베딩 벡터가 록업 테이블 형태로 저장된 임베딩 벡터 정보 테이블에서 상기 유전자 정보 중 1에 해당하는 유전자 정보의 임베딩 벡터를 추출하여 합산하며,

합산된 상기 유전자 정보를 이진화된 상기 임상 정보와 결합하여 기 구축된 치료 기간 예측 모델에 적용하여 2차 치료제에 대한 치료 기간을 각각 예측하는 2차 치료제 예후 예측 시스템.

#### 청구항 11

제7항에 있어서,  
 상기 예후 예측부는,  
 예측된 치료 기간이 제1 기간 미만인 경우 2차 치료제에 대한 반응 없음으로 예측하고,  
 예측된 치료 기간이 제1 기간 이상이고 제2 기간 미만이면 2차 치료제에 대한 반응 알 수 없음으로 예측하며,  
 예측된 치료 기간이 제2 기간 이상이면 2차 치료제에 대한 반응 있음으로 예측하는 2차 치료제 예후 예측 시스템.

## 청구항 12

제8항에 있어서,  
 상기 치료 기간 예측부는,  
 유전자에서 EZH2, FGFR1, CALR, TOP1, SRSF2, PIK3CA, ALK 중 적어도 하나에 유전적 변이가 있거나 조직학적 특성이 미만형인 경우 2차 치료제의 치료 기간을 제2 기간 이상으로 예측하고,  
 유전자에서 EGFR, SMARCA4, MTOR, APC, PBRM1, FBXW7, EGFR 중 적어도 하나에 유전적 변이가 있거나 초기에 암이 전이된 환자의 경우 2차 치료제의 치료 기간이 제1 기간 미만으로 예측하는 2차 치료제 예후 예측 시스템.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 위암 환자의 2차 치료제 예후 예측 방법 및 그 시스템에 관한 것으로서, 더욱 상세하게 설명하면 대상자의 임상 정보 및 유전자 정보를 고려하여 인공지능 모델을 통해 하나 이상의 2차 치료제의 예후를 각각 예측하는 위암 환자의 2차 치료제 예후 예측 방법 및 그 시스템에 관한 것이다.

### 배경 기술

[0002] 위암은 위에서 발생하는 악성 종양으로, 조기 위암의 경우 내시경 치료를 시행하지만, 2~3기의 진행성 위암은 수술로 암과 림프절을 제거하며, 암이 더욱 진행되어 위 주변 림프절이 아닌 멀리 떨어진 곳의 림프절에까지 전이되거나 복막에 파종되거나 다른 장기로 전이된 경우 항암화학요법을 통해 치료를 수행하게 된다.

[0003] 위암의 항암화학요법에 이용되는 1차 약물인 티에스원(Tegafur, TS-1), 카페시타빈(Capecitabine), 5-플루오로유라실(Fluorouracil, 5-FU) 중 하나와 옥살리플라틴(Oxaliplatin)이나 시스플라틴(Cisplatin)을 병합하여 사용한다. 1차 약물이 효과가 없는 경우 2차 약제로 탁산(Taxane) 계열의 치료제와 이리노테칸(Irinotecan) 계열의 치료제 중 어느 하나를 사용하고 3차 약제로 탁산 계열의 치료제와 이리노테칸 계열의 치료제 중 나머지 하나를 사용한다. 그러나 환자 개인의 임상적 요인 또는 유전적 요인에 따라 치료제에 대한 반응이 상이하므로 환자 개개인에 적합한 치료제를 선택할 수 있도록 하기 위해 각 치료제의 효과를 사전에 예측하는 기술이 필요한 실정이다.

## 발명의 내용

### 해결하려는 과제

[0004] 이와 같이 본 발명에 따르면, 대상자의 임상 정보 및 유전자 정보를 고려하여 인공지능 모델을 통해 하나 이상의 2차 치료제의 예후를 각각 예측하는 위암 환자의 2차 치료제 예후 예측 방법 및 그 시스템을 제공하기 위한 것이다.

### 과제의 해결 수단

[0005] 이러한 기술적 과제를 이루기 위한 본 발명의 실시예에 따르면 2차 치료제 예후 예측 방법에 있어서, 대상자 데이터 입력부가 대상자의 임상 정보 및 유전자 정보를 포함하는 대상자 데이터를 입력받는 단계; 치료 기간 예측부가 입력된 상기 대상자 데이터를 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료

제 각각에 대한 치료 기간을 예측하는 단계; 및 예후 예측부가 예측된 상기 치료 기간에 따라 대상자의 2차 치료제에 대한 반응 및 예후를 예측하는 단계를 포함할 수 있다.

[0006] 상기 임상 정보는, 대상자의 성별, 나이, 조직학적 특성, 전이 및 재발 여부, 1차 치료에서 표적치료제 사용 여부 및 1차 치료제 치료 기간을 포함하고,

[0007] 상기 유전자 정보는, 대상자의 신체 일 부위의 암에 대한 병원성 유전자의 단일 염기서열 변이 및 유전자 복제 변이 유무와 발암성 유전자의 단일 염기서열 변이 및 유전자 복제 변이 유무에 대한 정보를 포함할 수 있다.

[0008] 상기 치료 기간을 예측하는 단계는, 입력된 대상자 데이터 중 유전자 정보에 기 설정된 임베딩 방법을 통해 벡터 형태로 인코딩하고 상기 임상 정보와 임베딩된 상기 유전자 정보를 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료제 각각에 대한 치료 기간을 병렬적으로 예측할 수 있다.

[0009] 상기 치료 기간을 예측하는 단계는, 상기 치료 기간을 예측하는 단계는, 상기 임상 정보 및 유전자 정보를 각각 0 또는 1로 이진화하고, 모든 유전자에 대한 임베딩 벡터가 룩업 테이블 형태로 저장된 임베딩 벡터 정보 테이블에서 상기 유전자 정보 중 1에 해당하는 유전자 정보의 임베딩 벡터를 추출하여 합산하며, 합산된 상기 유전자 정보를 이진화된 상기 임상 정보와 결합하여 기 구축된 치료 기간 예측 모델에 적용하여 2차 치료제에 대한 치료 기간을 각각 예측할 수 있다.

[0010] 상기 반응 및 예후를 예측하는 단계는, 예측된 치료 기간이 제1 기간 미만인 경우 2차 치료제에 대한 반응 없음으로 예측하고, 예측된 치료 기간이 제1 기간 이상이고 제2 기간 미만이면 2차 치료제에 대한 반응 알 수 없음으로 예측하며, 예측된 치료 기간이 제2 기간 이상이면 2차 치료제에 대한 반응 있음으로 예측할 수 있다.

[0011] 상기 치료 기간을 예측하는 단계는, 유전자에서 EZH2, FGFR1, CALR, TOP1, SRSF2, PIK3CA, ALK 중 적어도 하나에 유전적 변이가 있거나 조직학적 특성이 미만형인 경우 2차 치료제의 치료 기간을 제2 기간 이상으로 예측하고, 유전자에서 EGFR, SMARCA4, MTOR, APC, PBRM1, FBXW7, EGFR 중 적어도 하나에 유전적 변이가 있거나 초기에 암이 전이된 환자의 경우 2차 치료제의 치료 기간이 제1 기간 미만으로 예측할 수 있다.

[0012] 본 발명의 다른 실시예에 따르면 2차 치료제 예후 예측 시스템에 있어서, 대상자의 임상 정보 및 유전자 정보를 포함하는 대상자 데이터를 입력받는 대상자 데이터 입력부; 입력된 상기 대상자 데이터를 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료제 각각에 대한 치료 기간을 예측하는 치료 기간 예측부; 및 예측된 상기 치료 기간에 따라 대상자의 2차 치료제에 대한 반응 및 예후를 예측하는 예후 예측부를 포함할 수 있다.

### 발명의 효과

[0013] 이와 같이 본 발명에 따르면, 2차 치료제에 대한 환자의 치료 기간을 예측하여 체계적으로 환자의 치료 계획을 수립하는데 활용될 수 있다.

[0014] 또한, 종래의 항암화학요법에 대한 반응성을 예측하기 위한 방법들은 모든 약물의 수의 제공에 대하여 실험이 요구되므로 시간, 비용 및 노동력 측면에 있어서 매우 비효율적이었으나 대상자의 임상 정보 및 유전자 정보를 기 구축된 인공지능 모델에 적용하여 치료 기간을 예측함으로써 신속하고 간편하게 2차 치료제 치료에 대한 반응 및 예후를 신뢰성있게 예측할 수 있다.

### 도면의 간단한 설명

[0015] 도 1은 본 발명의 일 실시예에 따른 2차 치료제 예후 예측 시스템을 도시하는 도면이다.

도 2는 본 발명의 다른 실시예에 따른 2차 치료제 예후 예측 방법을 나타내는 순서도이다.

도 3은 본 발명의 다른 실시예에 따른 임상 정보의 일 예를 도시하는 도면이다.

도 4는 본 발명의 다른 실시예에 따른 임상 정보 분포의 일 예를 도시하는 도면이다.

도 5는 본 발명의 다른 실시예에 따른 치료 기간 예측 모델의 구조를 도시하는 도면이다.

도 6은 본 발명의 다른 실시예에 따른 인공지능 모델에 입력되는 입력 데이터를 도시하는 도면이다.

도 7은 본 발명의 다른 실시예에 따른 환자의 상태를 예측하는데 높은 영향을 준 변수를 도시하는 도면이다.

도 8은 종래의 모델과 본 발명의 다른 실시예에 따른 치료 기간 예측 모델의 정확도를 비교하는 도면이다.

도 9는 종래의 모델과 본 발명의 다른 실시예에 따른 치료 기간 예측 모델의 생존곡선을 비교하는 도면이다.

### 발명을 실시하기 위한 구체적인 내용

- [0016] 이하 첨부된 도면을 참조하여 본 발명에 따른 바람직한 실시예를 상세히 설명하기로 한다. 이 과정에서 도면에 도시된 선들의 두께나 구성요소의 크기 등은 설명의 명료성과 편의상 과장되게 도시되어 있을 수 있다.
- [0017] 명세서 전체에서, 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.
- [0018] 또한 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서, 이는 사용자, 운용자의 의도 또는 관례에 따라 달라질 수 있다. 그러므로 이러한 용어들에 대한 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [0019] 후술하는 본 발명의 실시예에서 2차 치료제 예후 예측 시스템(100)은 위암에 대한 2차 치료제 예후를 예측하는 것으로 구체적인 예를 들어 설명한다. 그러나, 본 발명이 이로 제한되는 것은 아니다.
- [0020] 도 1은 본 발명의 일 실시예에 따른 2차 치료제 예후 예측 시스템을 도시하는 도면이다.
- [0021] 도 1에 도시된 바와 같이 2차 치료제 예후 예측 시스템(100)은 대상자 데이터 입력부(110), 치료 기간 예측부(120) 및 예후 예측부(130)를 포함하여 구성될 수 있다.
- [0022] 먼저, 대상자 데이터 입력부(110)는 대상자의 임상 정보 및 유전자 정보를 포함하는 대상자 데이터를 입력받을 수 있다.
- [0023] 여기서, 임상 정보는 대상자의 성별, 나이, 조직학적 특성(예를 들어, 장형(Intestinal), 미만형(Diffuse)), 전이 및 재발 여부, 1차 치료에서 표적치료제 사용 여부(예를 들어, 트라스투주맙(Trastuzumab)) 및 1차 치료제 치료 기간 중 적어도 하나를 포함하고, 유전자 정보는 대상자의 위암에 대한 병원성(Pathogenic) 또는 발암성(Oncogenic) 유전자의 단일 염기서열 변이(Single nucleotide variation, SNV) 및 유전자 복제 변이(Copy number variation, CNV) 유무에 대한 정보를 포함할 수 있다.
- [0024] 구체적으로, 대상자 데이터 입력부(110)는 전자건강기록(Electronic medical record, EMR), 병원정보시스템(Hospital information system, HIS), 임상병리검사기기, 디지털병리 시스템 및 데이터베이스(예를 들어, COSMIC, ClinVar, OncoKB) 중 적어도 하나로부터 대상자의 임상 정보 및 유전자 정보를 포함하는 대상자 데이터를 입력받을 수 있다.
- [0025] 다음으로, 치료 기간 예측부(120)는 입력된 대상자 데이터를 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료제 각각에 대한 치료 기간을 예측할 수 있다. 이때, 2차 치료제는 탁산(Taxane) 계열의 치료제(예를 들어, Paclitaxel 등) 또는 이리노테칸(Irinotecan) 계열의 치료제이고, 치료 기간 예측 모델은 각 2차 치료제에 대해 병렬적으로 치료 기간을 예측하도록 구축될 수 있다.
- [0026] 구체적으로, 치료 기간 예측부(120)는 입력된 대상자 데이터 중 유전자 정보에 기 설정된 임베딩 방법(예를 들어, 단어 임베딩(Word embedding))을 통해 벡터 형태로 인코딩하고 임상 정보와 임베딩된 유전자 정보를 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료제 각각에 대한 치료 기간을 예측할 수 있다.
- [0027] 여기서, 치료 기간 예측 모델을 이용하여 치료 기간을 예측하는 방법은 아래의 도 4 내지 도 8을 통해 보다 구체적으로 설명한다.
- [0028] 다음으로, 예후 예측부(130)는 예측된 치료 기간에 따라 대상자의 각 2차 치료제에 대한 반응 및 예후를 예측할 수 있다.
- [0029] 구체적으로, 예후 예측부(130)는 예측된 치료 기간이 기 설정된 제1 기간(예를 들면, 3개월)보다 짧은 경우 대상자의 2차 치료제에 대한 반응 없음으로 예측할 수 있다.
- [0030] 또한, 예후 예측부(130)는 예측된 치료 기간이 기 설정된 제2 기간(예를 들면, 6개월) 이상인 경우 대상자의 2차 치료제에 대한 반응 있음으로 예측할 수 있다. 이때, 예후 예측부(130)는 대상자의 예후를 긍정적으로 예측할 수 있다.
- [0031] 또한, 예후 예측부(130)는 예측된 치료 기간이 제1 기간(예를 들어, 3개월) 이상이고, 제2 기간(예를 들어, 6개

월)보다 짧으면 대상자의 2차 치료제에 대한 반응 알 수 없음으로 예측할 수 있다.

- [0032] 이하에서는 도 2 내지 도 8을 참조하여, 위암 환자의 2차 치료제 예후 예측 방법에 대해 보다 구체적으로 설명한다.
- [0033] 도 2는 본 발명의 다른 실시예에 따른 2차 치료제 예후 예측 방법을 나타내는 순서도이다.
- [0034] 먼저, 대상자 데이터 입력부(110)는 대상자의 임상 정보 및 유전자 정보를 포함하는 대상자 데이터를 입력받을 수 있다(S210).
- [0035] 구체적으로, 대상자 데이터 입력부(110)는 전자건강기록, 병원정보시스템, 임상병리검사기기, 디지털병리 시스템 및 데이터베이스(예를 들어, COSMIC, ClinVar, OncoKB 등) 중 적어도 하나로부터 대상자의 임상 정보 및 유전자 정보를 포함하는 대상자 데이터를 입력받을 수 있다.
- [0036] 여기서, 임상 정보는 대상자의 성별, 나이, 조직학적 특성(예를 들어, 장형 또는 미만형), 전이 및 재발 여부, 1차 치료에서 표적치료제 사용 여부(예를 들어, 트라스투주맙) 및 1차 치료제 치료 기간 중 적어도 하나를 포함할 수 있다.
- [0037] 도 3은 본 발명의 다른 실시예에 따른 임상 정보의 일 예를 도시하는 도면이다.
- [0038] 도 3에 도시된 바와 같이, 임상 정보는 대상자의 등록 번호(S번호), 나이(Age), 성별(Sex), 조직학적 유형(Histology), HER2 상태(HER2 status), 초기 환경(Initial setting), 치료 위절제술(Curative gastrectomy) 시행 여부, 보조 항암화학요법(Adjuvant CTx) 또는 선행 항암화학요법(Neoadjuvant CTx) 시행 여부 및 확인 사항, 1차 치료제 치료 기간(1L chemotherapy)일 수 있다.
- [0039] 또한, 유전자 정보는 대상자의 위암에 대한 병원성 또는 발암성 유전자의 단일 염기서열 변이 및 유전자 복제 변이 유무에 대한 정보를 포함할 수 있다.
- [0040] 다음으로, 치료 기간 예측부(120)는 입력된 대상자 데이터를 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료제 각각에 대한 치료 기간을 예측할 수 있다(S220). 이때, 2차 치료제는 탁산 계열의 치료제(예를 들어, Paclitaxel 등) 또는 이리노테칸 계열의 치료제이고, 치료 기간 예측 모델은 각 2차 치료제에 대해 병렬적으로 치료 기간을 예측하도록 구축될 수 있다.
- [0041] 구체적으로, 치료 기간 예측부(120)는 입력된 대상자 데이터 중 유전자 정보에 기 설정된 임베딩 방법(예를 들어, 단어 임베딩)을 적용하여 벡터 형태로 인코딩하고 이를 임상 정보와 함께 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료제 각각에 대한 치료 기간을 예측하도록 구축될 수 있다.
- [0042] 여기서, 치료 기간 예측 모델은 임베딩된 암 종류별 병원성 유전자의 단일 염기서열 변이 및 유전자 복제 변이 유무, 발암성 유전자의 단일 염기서열 변이 및 유전자 복제 변이 유무와 성별, 나이, 조직학적 특성(예를 들어, 장형 또는 미만형), 전이 및 재발 여부, 1차 치료에서 표적치료제 사용 여부(예를 들어, 트라스투주맙) 및 1차 치료제 치료 기간 및 2차 치료제 치료 기간을 포함하는 오픈소스 이중데이터를 학습 데이터셋으로 이용하여 인공 신경망(Artificial neural network, ANN)을 통해 하나 이상의 2차 치료제 치료 기간을 병렬적으로 예측하도록 구축될 수 있다.
- [0043] 다시 말해, 치료 기간 예측부(120)는 입력된 대상자 데이터 중 유전자 정보에 기 설정된 임베딩 방법을 통해 벡터 형태로 인코딩하고, 임상 정보와 임베딩된 유전자 정보를 기 구축된 치료 기간 예측 모델에 적용하여 기 설정된 하나 이상의 2차 치료제(예를 들어, 탁산 계열 치료제 및 이리노테칸 계열의 치료제)에 대한 치료 기간을 병렬적으로 예측할 수 있다.
- [0044] 또한, 치료 기간 예측부(120)는 6개월 이상 2차 치료제를 사용하는 위암 환자와 2차 치료제를 3개월 미만 사용하고 3차 치료제를 6개월 이상 사용하는 위암 환자의 임상 정보와 유전자 정보를 포함하는 학습 데이터로 하여 기 설정된 방법(예를 들어, 단어 임베딩)을 통해 임베딩을 수행하여 벡터 형태로 인코딩하고 이를 입력 데이터로 치료 기간 예측 모델에 적용하여 각 치료제에 대한 치료 기간을 출력하도록 학습할 수 있다.
- [0045] 즉, 치료 기간 예측부(120)는 임상 정보 및 유전자 정보를 각각 0 또는 1로 이진화하고, 모든 유전자에 대한 임베딩 벡터가 록업 테이블 형태로 저장된 임베딩 벡터 정보 테이블에서 유전자 정보 중 1에 해당하는 유전자 정보의 임베딩 벡터를 추출하여 합산하며, 합산된 유전자 정보를 이진화된 임상 정보와 결합하여 기 구축된 치료 기간 예측 모델에 적용하여 2차 치료제에 대한 치료 기간을 각각 예측할 수 있다.

- [0046] 다음으로, 예후 예측부(130)는 예측된 치료 기간에 따라 대상자의 각 치료제에 대한 반응 및 예후를 예측할 수 있다(S230).
- [0047] 구체적으로, 예후 예측부(130)는 예측된 치료 기간이 제1 기간(예를 들어, 3개월) 미만인 2차 치료제에 대해 반응 없음으로 예후를 예측할 수 있다.
- [0048] 또한, 예후 예측부(130)는 예측된 치료 기간이 제2 기간(예를 들어, 6개월) 이상인 2차 치료제에 대해 반응 있음으로 예후를 예측할 수 있다.
- [0049] 다시 말해, 특정 2차 치료제에 대해 치료 기간이 짧다는 것은 해당 2차 치료제의 효과가 떨어져서 해당 2차 치료제를 단 기간만 사용하고 다음 2차 치료제로 변경하였음을 의미하는 것이므로, 예후 예측부(130)는 해당 2차 치료제에 대해 반응 없음으로 예측할 수 있다.
- [0050] 도 2 및 도 3을 참조하여 상술한 위암 환자의 항암자별 예후 예측 방법은 이에 대한 예측이 가능한 컴퓨팅 장치 및 소프트웨어에 의해 수행될 수 있다.
- [0051] 이하에서는 도 4 내지 도 8을 이용하여 본 발명의 실시예에 따른 2차 치료제 예후 예측 방법에 대하여 상세하게 설명한다. 그러나, 하기의 실시예는 본 발명의 내용을 예시하는 것일 뿐, 본 발명의 범위가 하기의 실시예에 한정되는 것은 아니다. 본 발명의 실시예는 당업계에서 평균적인 지식을 가진 자에게 본 발명을 보다 완전하게 설명하기 위해 제공되는 것이다.
- [0052] **실험방법**
- [0053] <1-1> 실험 대상 설정
- [0054] K-MASTER 프로젝트에 등록된 위암 환자 857명 중 2차 치료제로 탁산 계열의 치료제(예를 들어, 파클리탁셀(Paclitaxel)) 기반의 항암 치료를 처방받은 환자 402명 중 2차 치료 기간이 3개월 이상이고 6개월 미만인 환자 144명을 제외하였다.
- [0055] 또한, 2차 치료 기간이 3개월 미만인 환자를 탁산 계열의 치료제에 저항성이 있는 환자군(Paclitaxel-resistant)으로 분류하고, 2차 치료 기간이 6개월 이상인 환자를 탁산 계열의 치료제에 민감한 환자군(Paclitaxel-sensitive)으로 분류하였다.
- [0056] 또한, 탁산 계열의 치료제에 저항성이 있는 환자군 및 탁산 계열의 치료제에 민감한 환자군 중 80%는 학습 데이터셋(Training set)으로 분류하고 20%를 검증 데이터셋(Validation set)으로 분류하였다.
- [0057] <1-2> 임상 정보 및 유전자 정보
- [0058] 임상 정보는 환자의 나이, 성별, 조직학적 특성(장형, 미만형), 암 전이 및 재발 여부, 1차 치료에서의 트라스투주맵 처방 여부 및 1차 치료 기간을 포함하고, 유전자 정보는 데이터베이스(예를 들어, COSMIC, ClinVar, OncoKB 등)의 위암에 대한 병원성 또는 발암성 유전자의 단일 염기서열 변이 및 유전자 복제 변이 유무에 대한 정보를 포함할 수 있다.
- [0059] 도 4는 본 발명의 다른 실시예에 따른 임상 정보 분포의 일 예를 도시하는 도면이다.
- [0060] 도 4의 (a)에 도시된 바와 같이, 학습 데이터셋과 검증 데이터셋의 임상 정보 분포를 알 수 있다. 이때, 학습 데이터셋과 검증 데이터셋의 나이는 유사하고, 성별은 학습 데이터셋이 검증 데이터셋에 비해 남성이 8.8%p 더 많고, 조직학적 특성은 장형이 검증 데이터셋이 10.2%p 더 많고 미만형이 10.2%p 더 적음을 알 수 있다.
- [0061] 도 4의 (b)에 도시된 바와 같이, 탁산 계열의 치료제에 저항성이 있는 환자군(Paclitaxel-resistant)과 탁산 계열의 치료제에 민감한 환자군(Paclitaxel-sensitive)의 임상 정보 분포를 알 수 있다. 이때, 탁산 계열의 치료제에 민감한 환자군이 탁산 계열의 치료제에 저항성이 있는 환자군보다 남성이 10.2%p 더 많고, 조직학적 특성에서 장형이 12.4%p 더 많고 민감형이 12.4%p 더 적음을 알 수 있다.
- [0062] 본 발명의 일 실시예에 따르면 데이터베이스(예를 들어, COSMIC, ClinVar, OncoKB)로부터 아래의 [표 1]에 도시된 87개의 유전자 정보를 사용하였다.

표 1

|        |       |         |        |       |        |       |        |        |
|--------|-------|---------|--------|-------|--------|-------|--------|--------|
| TP53   | APC   | BARD1   | TET2   | BCOR  | FANCA  | MLH1  | ERCC2  | RIT1   |
| PIK3CA | KRAS  | POLE    | STK11  | BRAF  | DNMT3A | FGFR1 | ERBB4  | CTNNB1 |
| CDH1   | SMAD4 | MSH6    | CALR   | MDM2  | PRDM1  | STAT3 | SDHB   | ESR1   |
| BRCA2  | PTCH1 | PTPRD   | ASXL1  | CSF3R | GNA11  | KIT   | RICTOR | SF3B1  |
| BRCA1  | ATM   | CDKN2A  | TERT   | RB1   | CDK12  | NCOA3 | FBXW7  | NKX2-1 |
| ERBB2  | NF1   | EGFR    | KMT2D  | TOP1  | CDK4   | GNAS  | GNAQ   | MAP2K1 |
| MET    | PTEN  | PTPN11  | ARID1A | ERBB3 | NRAS   | IDH2  | WT1    | CEBPA  |
| FGFR3  | SRSF2 | SMARCA4 | CCND1  | HRAS  | ALK    | MAPK1 | MYC    |        |
| AR     | CCNE1 | PBRM1   | MTOR   | ASNS  | SMO    | MRE11 | PIK3R1 |        |
| NOTCH1 | FGFR2 | FLT3    | FGF3   | RUNX1 | EZH2   | MSH2  | AXIN2  |        |

[0063]

[0064]

이때, 많이 발견된 유전자는 TP53, PIK3CA, CDH1, BRCA1, BRCA2, ERBB2, MET, FGFR3 등 기존 위암에서 중요하게 발견된 유전자로 구성되어 있고, Paclitaxel 약물 치료와 관련하여 유의미한 차이를 보인 유전자는 TET2, PTPN11, CSF3R, CALR, TOP1, EGFR이다.

[0065]

한편, 이러한 유전자 변이들이 복합적으로 작용하여 Paclitaxel 치료제의 반응에 영향을 끼치는 것을 알 수 있다.

[0066]

### <1-3> 치료 기간 예측 모델 비교

[0067]

임상 정보와 유전자 정보를 입력 데이터로 하는 랜덤 포레스트(Random forest, RF), 회귀 분석(Logistic regression, LR) 및 인공 신경망과 본 발명의 실시예에 따라 유전자 정보를 임베딩하고 임상 정보와 함께 입력 데이터로 하는 치료 기간 예측 모델을 2차 치료제에 대한 치료 기간을 병렬적으로 예측하도록 각 모델을 학습하였다.

[0068]

도 5는 본 발명의 다른 실시예에 따른 치료 기간 예측 모델의 구조를 도시하는 도면이다.

[0069]

도 5에 도시된 바와 같이, 치료 기간 예측 모델은 임상 정보와 임베딩된 유전자 정보를 입력 데이터로 하는 인공 신경망을 기초로 구축된 모델로 2차 치료제에 대한 치료 기간을 예측할 수 있다.

[0070]

도 6은 본 발명의 다른 실시예에 따른 인공지능 모델에 입력되는 입력 데이터를 도시하는 도면이다.

[0071]

도 6에 도시된 바와 같이, 종래의 모델(랜덤 포레스트, 회귀 분석, 인공 신경망)에 입력되는 임상 정보(Clinical feature) 및 유전자 정보(Genetic feature)는 각각 0 또는 1로 이진화되어 입력된다.

[0072]

또한, 치료 기간 예측 모델에 입력되는 임상 정보 및 유전자 정보는 각각 0 또는 1로 이진화된다.

[0073]

그 다음, 모든 유전자에 대한 임베딩 벡터가 룩업 테이블 형태로 저장되어 있는 임베딩 벡터 정보 테이블(All genetic variant embedding lookup table)에서 '1'에 해당하는 이진화된 유전자 정보의 임베딩 벡터를 추출하여 합산(Sum of variant embeddings)한다.

[0074]

그 다음, 합산된 유전자 정보를 이진화된 임상 정보와 결합하여(Concatenated clinical and genetic vectors) 치료 기간 예측 모델에 입력한다.

[0075]

### 실험결과

[0076]

### <2-1> 예측 결과

[0077]

도 7은 본 발명의 다른 실시예에 따른 환자의 상태를 예측하는데 높은 영향을 준 변수를 도시하는 도면이다.

[0078]

도 7에 도시된 바와 같이, 2차 치료제에 대해 반응 있음으로 예측하도록 영향을 준 변수는 EZH2, FGFR1, CALR, TOP1, Histology-Diffuse type, SRSF2, PIK3CA, ALK 등 으로, 해당 유전자에 유전적 변이가 있거나 조직학적 특성이 미만형인 경우 2차 치료제의 치료 기간이 제2 기간 이상으로 예측됐다.

[0079]

또한, 2차 치료제에 대해 반응 없음으로 예측하도록 영향을 준 변수는 EGFR, Initial setting-Initially metastatic, SMARCA4, MTOR, APC, PBRM1, FBXW7, EGFR 등 으로, 해당 유전자에 유전적 변이가 있거나 초기에 위암이 전이된 환자의 경우 2차 치료제의 치료 기간이 제1 기간 미만으로 예측됐다.

[0080]

또한, 제1 기간 미만 또는 제2 기간 이상을 예측하기에 높은 영향을 기여한 변수의 순서대로 가중치를 추출하여 치료기간 예측 모델에 적용하여 정확도를 향상시킬 수 있다.

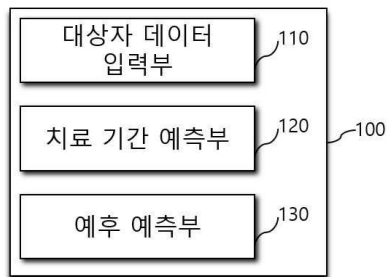
- [0081] 도 8은 종래의 모델과 본 발명의 다른 실시예에 따른 치료 기간 예측 모델의 정확도를 비교하는 도면이다.
- [0082] 도 8에 도시된 바와 같이, 정확도(Accuracy)는 종래의 모델 중 인공신경망(ANN)이 0.621인 반면, 본 발명의 다른 실시예에 따른 치료 기간 예측 모델(ANN with GE)은 0.724로 더 정확하게 예측함을 알 수 있다. 이때, 정확도는 전체 샘플 중에서 올바르게 예측한 샘플의 비율이다.
- [0083] 또한, 민감도(Sensitivity)는 종래의 모델 중 인공신경망(ANN)이 0.500인 반면, 본 발명의 다른 실시예에 따른 치료 기간 예측 모델(ANN with GE)은 0.458로 유사함을 알 수 있다. 이때, 민감도는 실제 양성 샘플 중에서 올바르게 예측한 양성 샘플의 비율이다.
- [0084] 또한, 특이도(Specificity)는 종래의 모델 중 인공신경망(ANN)이 0.706인 반면, 본 발명의 다른 실시예에 따른 치료 기간 예측 모델(ANN with GE)은 0.912로 더 올바르게 예측함을 알 수 있다. 이때, 특이도는 실제 음성 샘플 중에서 올바르게 예측한 음성 샘플의 비율이다.
- [0085] 또한, F1 점수(F1 score)는 종래의 모델(RF, LR, ANN)이 0.417 내지 0.522 인 반면 본 발명의 다른 실시예에 따른 치료 기간 예측 모델(ANN with GE)은 0.579로 분류 성능이 우수함을 알 수 있다.
- [0086] 또한, 수신자 조작 특성(Area under ROC, AUROC)은 종래의 모델(RF, LR, ANN)이 0.499 내지 0.679 인 반면 본 발명의 다른 실시예에 따른 치료 기간 예측 모델(ANN with GE)은 0.732로 예측 성능이 가장 우수함을 알 수 있다.
- [0087] 도 9는 종래의 모델과 본 발명의 다른 실시예에 따른 치료 기간 예측 모델의 생존곡선을 비교하는 도면이다.
- [0088] 도 9의 (a) 내지 (h)에 도시된 바와 같이, 종래의 모델(Random forest, Logistic Regression, Artificial neural network)과 치료 기간 예측 모델(Artificial neural network with GE)의 검증 데이터셋에 대한 예측 그룹의 치료 기간(Progression-free survival, PFS)(도 9의 (a) 내지 (d))과 실제 생존 기간(도 9의 (e) 내지 (h))을 비교하면 치료 기간 예측 모델이 종래의 모델에 비해 더 정확하게 타산 계열의 치료제에 저항성이 있는 환자군(Predicted Pac-resistant)(빨간색 선)과 타산 계열의 치료제에 민감한 환자군(Predicted Pac-sensitive)(초록색 선)을 분류함을 알 수 있다.
- [0089] 본 발명의 실시예에 따르면, 2차 치료제에 대한 환자의 치료 기간을 예측하여 체계적으로 환자의 치료 계획을 수립하는데 활용될 수 있다.
- [0090] 또한, 종래의 항암화학요법에 대한 반응성을 예측하기 위한 방법들은 모든 약물의 수의 제공에 대하여 실험이 요구되므로 시간, 비용 및 노동력 측면에 있어서 매우 비효율적이었으나 대상자의 임상 정보 및 유전자 정보를 기 구축된 인공지능 모델에 적용하여 치료 기간을 예측함으로써 신속하고 간편하게 2차 치료제 치료에 대한 반응 및 예후를 신뢰성있게 예측할 수 있다.
- [0091] 본 발명은 도면에 도시된 실시예를 참고로 하여 설명되었으나 이는 예시적인 것에 불과하며, 당해 기술이 속하는 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서 본 발명의 진정한 기술적 보호범위는 아래의 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

## 부호의 설명

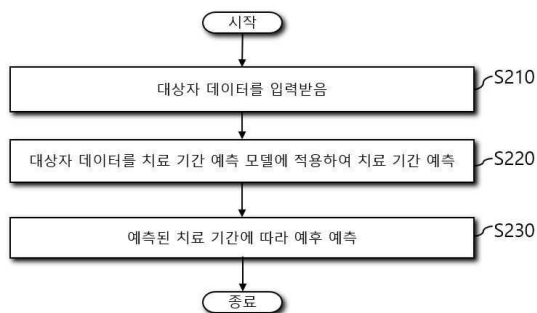
- [0092] 100: 2차 치료제 예후 예측 시스템  
110: 대상자 데이터 입력부  
120: 치료 기간 예측부  
130: 예후 예측부

## 도면

도면1



도면2



도면3

| No. | ID     | Age | Sex | Histology         | MSI status | Initial setting    | Current<br>pathologic<br>stage | Neoadjuvant or adjuvant CTx regimen |                         |                       |                       | PD for<br>adjuvant<br>monotherapy | PD for<br>adjuvant<br>monotherapy | S1 chemotherapy |                        |                       |                     | S2 chemotherapy    |                        |                      |                       | Death                 | Relapse |                    |                 |       |
|-----|--------|-----|-----|-------------------|------------|--------------------|--------------------------------|-------------------------------------|-------------------------|-----------------------|-----------------------|-----------------------------------|-----------------------------------|-----------------|------------------------|-----------------------|---------------------|--------------------|------------------------|----------------------|-----------------------|-----------------------|---------|--------------------|-----------------|-------|
|     |        |     |     |                   |            |                    |                                | Neoadjuvant CTx<br>regimen          | Adjuvant CTx<br>regimen | Adjuvant CTx<br>stage | Adjuvant CTx<br>stage |                                   |                                   | S1 CTx type     | S1 CTx regimen details | S1 CTx start<br>date  | S1 CTx last<br>date | S1 CTx<br>duration | S1 CTx regimen details | S1 CTx start<br>date | S1 CTx last<br>date   |                       |         | S1 CTx<br>duration | S1 CTx<br>stage |       |
| 1   | 510124 | 70  | 남   | Endometrioid type | Negative   | Recurrent disease  | 1                              | 0                                   |                         |                       |                       |                                   |                                   | 0               | 0                      | FOU/FOX/VEL/OL/MP/MTX | Pembrolizumab+MSKCA | 2020/10/08         | 2020/11/27             | 9.83                 |                       | FOU/FOX/VEL/OL/MP/MTX | 15.91   | 1                  | 0               | 19.10 |
| 2   | 510089 | 68  | 여   | Diffuse type      | Negative   | Recurrent disease  | 1                              | 1                                   | MSKCA                   | 2020/01/03            | 2020/12/27            | 4.83                              | 1                                 | 0               | Adjuvant_MSKCA         | Adjuvant_MSKCA        | 2020/01/03          | 2020/12/27         | 8.83                   | 1                    | FOU/FOX/VEL/OL/MP/MTX | 8.52                  | 1       | 0                  | 14.70           |       |
| 3   | 510059 | 58  | 남   | Diffuse type      | Negative   | metastatic disease | 0                              | 0                                   |                         |                       |                       |                                   |                                   | 0               | 0                      | FOU/FOX/VEL/OL/MP/MTX | Teladomab+MSKCA     | 2020/02/14         | 2020/11/09             | 8.52                 |                       | FOU/FOX/VEL/OL/MP/MTX | 5.06    | 1                  | 1               | 7.82  |
| 4   | 509929 | 77  | 남   | Diffuse type      | Negative   | metastatic disease | 1                              | 0                                   |                         |                       |                       |                                   |                                   | 0               | 0                      | MSKCA                 | 2020/09/05          | 2020/12/24         | 8.75                   | 1                    | Remotumab+FOU/FOX     | 8.84                  | 0       | 0                  | 9.94            |       |
| 5   | 508524 | 70  | 남   | Unknown           | Unknown    | metastatic disease | 0                              | 0                                   |                         |                       |                       |                                   |                                   | 0               | 0                      | MSKCA                 | 2020/11/26          | 2021/06/08         | 6.38                   | 1                    | Remotumab+FOU/FOX     | 0.49                  | 1       | 1                  | 0.44            |       |
| 7   | 508466 | 75  | 남   | Unknown           | Positive   | metastatic disease | 0                              | 0                                   |                         |                       |                       |                                   |                                   | 0               | 1                      | HFP                   | 2020/03/24          | 2020/12/01         | 8.28                   | 1                    | Remotumab+FOU/FOX     | 2.07                  | 1       | 0                  | 15.10           |       |
| 8   | 505981 | 66  | 여   | Unknown           | Negative   | metastatic disease | 0                              | 0                                   |                         |                       |                       |                                   |                                   | 0               | 0                      | SP                    | 2020/11/19          | 2020/06/01         | 6.87                   | 1                    | Remotumab+FOU/FOX     | 3.52                  | 1       | 0                  | 7.64            |       |
| 11  | 505871 | 79  | 여   | Unknown           | Negative   | metastatic disease | 0                              | 0                                   |                         |                       |                       |                                   |                                   | 0               | 0                      | SP                    | 2020/12/18          | 2020/01/08         | 6.74                   | 1                    | Vincoreb+FOU/FOX      | 6.15                  | 1       | 1                  | 9.17            |       |
| 13  | 505801 | 41  | 여   | Unknown           | Negative   | metastatic disease | 0                              | 0                                   |                         |                       |                       |                                   |                                   | 0               | 0                      | FOU/FOX               | 2020/05/14          | 2020/09/20         | 4.24                   | 1                    | Remotumab+FOU/FOX     | 5.23                  | 1       | 1                  | 12.11           |       |
| 14  | 505800 | 70  | 여   | Diffuse type      | Negative   | metastatic disease | 1                              | 0                                   |                         |                       |                       |                                   |                                   | 0               | 0                      | MSKCA                 | 2020/05/24          | 2020/10/16         | 4.77                   | 1                    | Remotumab+FOU/FOX     | 8.65                  | 1       | 1                  | 17.04           |       |
| 15  | 505767 | 74  | 여   | Mixed type        | Negative   | Recurrent disease  | 1                              | 0                                   |                         |                       |                       |                                   |                                   | 0               | 0                      | SP                    | 2020/04/18          | 2021/04/18         | 12.07                  | 1                    | Remotumab+FOU/FOX     | 4.64                  | 1       | 0                  | 11.34           |       |

도면4

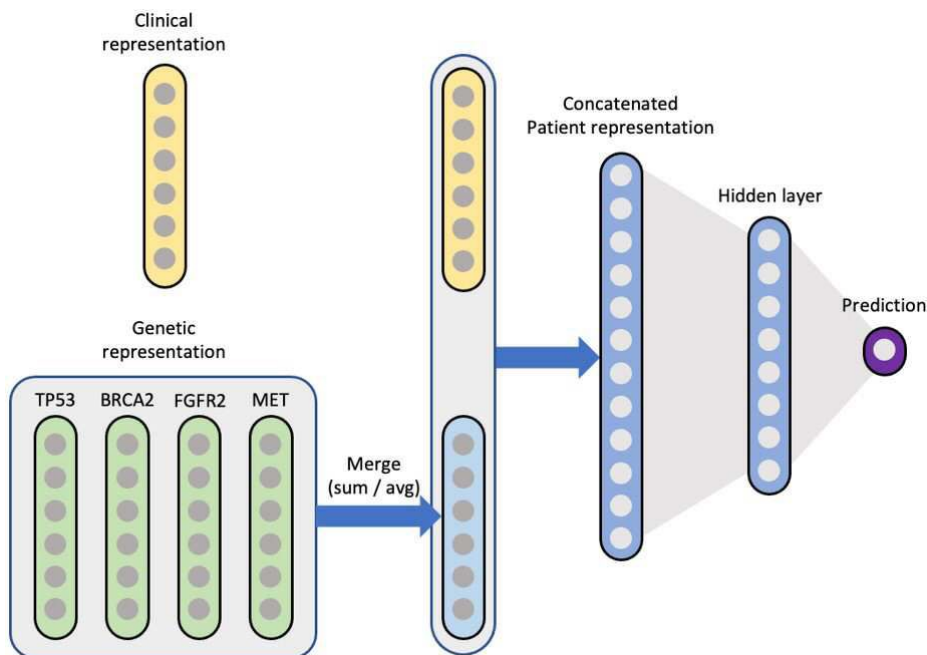
|                                                                                       | Training set<br>(N=230, %) | Validation<br>set (N=58,<br>%) | P     |                                                                                     | Paclitaxel-<br>sensitive<br>(N=93, %) | Paclitaxel-<br>resistant<br>(N=137, %) | P     |
|---------------------------------------------------------------------------------------|----------------------------|--------------------------------|-------|-------------------------------------------------------------------------------------|---------------------------------------|----------------------------------------|-------|
| Age (y)                                                                               |                            |                                |       | Age (y)                                                                             |                                       |                                        |       |
| Median, years (range)                                                                 | 64 (25-91)                 | 62 (28-89)                     | 0.154 | Median, years (range)                                                               | 64 (37-87)                            | 64 (25-91)                             | 0.910 |
| <40 years, n (%)                                                                      | 8 (3.5)                    | 5 (8.6)                        | 0.153 | <40 years, n (%)                                                                    | 3 (3.2)                               | 5 (3.6)                                | 0.833 |
| ≥ 70 years, n (%)                                                                     | 74 (32.2)                  | 14 (24.1)                      |       | ≥ 70 years, n (%)                                                                   | 28 (30.1)                             | 46 (33.6)                              |       |
| Sex                                                                                   |                            |                                | 0.209 | Sex                                                                                 |                                       |                                        | 0.116 |
| Male                                                                                  | 155 (67.4)                 | 34 (58.6)                      |       | Male                                                                                | 57 (61.3)                             | 98 (71.5)                              |       |
| Female                                                                                | 75 (32.6)                  | 24 (41.4)                      |       | Female                                                                              | 36 (28.5)                             | 39 (28.5)                              |       |
| Histology                                                                             |                            |                                | 0.146 | Histology                                                                           |                                       |                                        | 0.072 |
| Intestinal type                                                                       | 143 (62.1)                 | 42 (72.4)                      |       | Intestinal type                                                                     | 51 (54.8)                             | 92 (67.2)                              |       |
| Diffuse type*                                                                         | 87 (37.8)                  | 16 (27.6)                      |       | Diffuse type                                                                        | 42 (45.2)                             | 45 (32.8)                              |       |
| Clinical settings                                                                     |                            |                                | 0.835 | Clinical settings                                                                   |                                       |                                        | 0.132 |
| Initially metastatic disease                                                          | 154 (67.0)                 | 38 (65.6)                      |       | Initially metastatic disease                                                        | 57 (61.3)                             | 97 (70.8)                              |       |
| Recurrent disease                                                                     | 76 (33.0)                  | 20 (34.5)                      |       | Recurrent disease                                                                   | 36 (38.7)                             | 40 (29.2)                              |       |
| Prior gastrectomy                                                                     | 120 (52.2)                 | 26 (44.8)                      | 0.353 | Prior gastrectomy                                                                   | 56 (60.2)                             | 64 (46.7)                              | 0.059 |
| Prior trastuzumab use                                                                 | 40 (17.4)                  | 8 (13.8)                       | 0.511 | Prior trastuzumab use                                                               | 20 (21.5)                             | 20 (14.6)                              | 0.215 |
| Duration of prior 1st-line fluoro-<br>pyrimidine plus platinum- based<br>chemotherapy |                            |                                | 0.508 | Duration of prior 1st-line<br>fluoropyrimidine plus platinum-<br>based chemotherapy |                                       |                                        | 0.034 |
| < 3 months                                                                            | 57 (24.8)                  | 16 (27.6)                      |       | < 3 months                                                                          | 17 (18.3)                             | 40 (29.2)                              |       |
| ≥ 6 months                                                                            | 87 (37.8)                  | 19 (32.7)                      |       | ≥ 6 months                                                                          | 44 (47.3)                             | 43 (31.4)                              |       |

\* Mixed type was categorized into diffuse type: training set (n=13) and validation set (n=1).

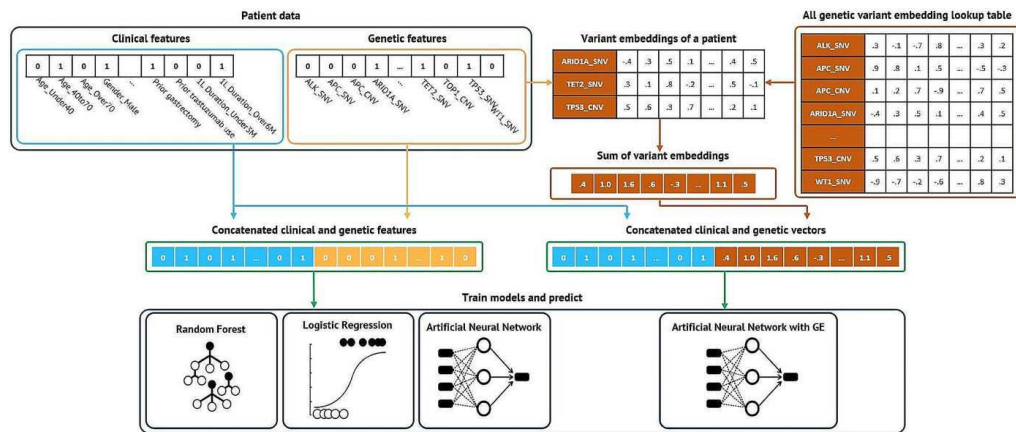
(a)

(b)

도면5



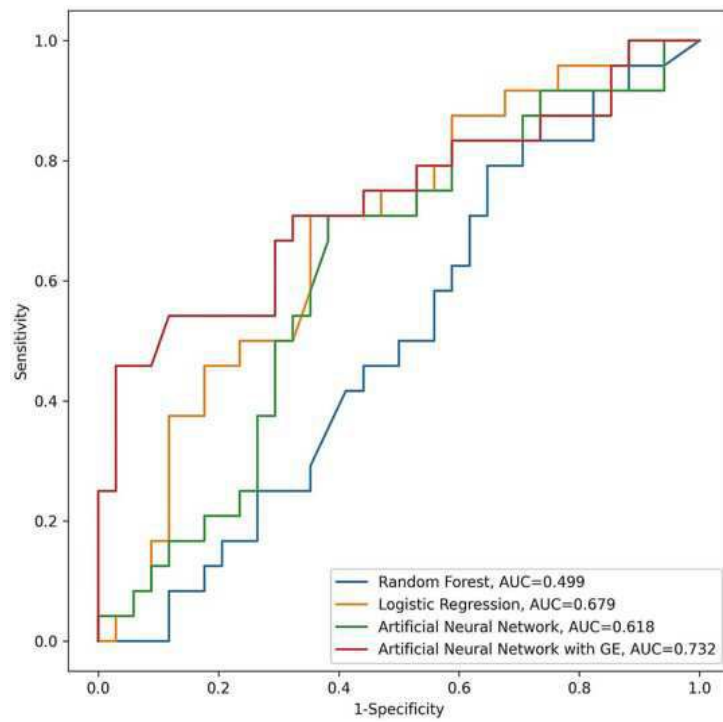
도면6



도면7

| Bottom-20 Coefficients |                                              |          | Top-20 Coefficients |                        |          |
|------------------------|----------------------------------------------|----------|---------------------|------------------------|----------|
|                        | Column                                       | Coef     |                     | Column                 | Coef     |
| 27                     | EGFR_SNV                                     | -0.96205 | 55                  | EZH2_SNV               | 0.436808 |
| 7                      | Initial setting-Initially metastatic disease | -0.90108 | 68                  | FGFR1_SNV              | 0.437009 |
| 23                     | SMARCA4_SNV                                  | -0.61386 | 79                  | CALR_SNV               | 0.44147  |
| 89                     | MTOR_SNV                                     | -0.59062 | 87                  | TOP1_CNV               | 0.441689 |
| 64                     | APC_SNV                                      | -0.52896 | 4                   | Histology-Diffuse type | 0.450829 |
| 82                     | PBRM1_SNV                                    | -0.50184 | 56                  | SRSF2_CNV              | 0.450953 |
| 6                      | Initial setting-Recurrent disease            | -0.49359 | 67                  | PIK3CA_CNV             | 0.458842 |
| 93                     | FBXW7_SNV                                    | -0.48651 | 37                  | ALK_SNV                | 0.464634 |
| 84                     | EGFR_CNV                                     | -0.4811  | 71                  | CALR_CNV               | 0.479657 |
| 101                    | RIT1_SNV                                     | -0.44607 | 10                  | 1L_Duration_6Mto       | 0.492094 |
| 80                     | ASNS_CNV                                     | -0.44302 | 40                  | POLE_SNV               | 0.509426 |
| 78                     | SDHB_SNV                                     | -0.43448 | 43                  | CSF3R_SNV              | 0.550462 |
| 20                     | ASXL1_SNV                                    | -0.42687 | 12                  | NRAS_SNV               | 0.557318 |
| 9                      | 1L_Duration_to3M                             | -0.39507 | 65                  | MLH1_SNV               | 0.561065 |
| 107                    | FANCA_SNV                                    | -0.37075 | 22                  | NF1_SNV                | 0.568747 |
| 63                     | ERCC2_SNV                                    | -0.36252 | 111                 | GNAS_SNV               | 0.615259 |
| 106                    | NKX2-1_SNV                                   | -0.35632 | 8                   | Prior trastuzumab use  | 0.63375  |
| 57                     | MSH2_SNV                                     | -0.336   | 110                 | MTOR_CNV               | 0.655061 |
| 30                     | AR_SNV                                       | -0.33372 | 49                  | TET2_SNV               | 0.890313 |
| 69                     | ERBB4_SNV                                    | -0.32877 | 48                  | PTPN11_SNV             | 1.19086  |

도면8



| Models         | AUROC | Sensitivity | Specificity | Accuracy | F1 score |
|----------------|-------|-------------|-------------|----------|----------|
| RF             | 0.499 | 0.517       | 0.417       | 0.588    | 0.417    |
| LR             | 0.679 | 0.638       | 0.375       | 0.823    | 0.461    |
| ANN            | 0.618 | 0.500       | 0.706       | 0.621    | 0.522    |
| ANN with<br>GE | 0.732 | 0.458       | 0.912       | 0.724    | 0.579    |

RF, random forest; LR, logistic regression; ANN, artificial neural network; ANN with GE, artificial neural network with genetic embedding

도면9

