



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0111941
(43) 공개일자 2020년10월05일

(51) 국제특허분류(Int. Cl.)

G16H 50/70 (2018.01) G01N 33/53 (2006.01)

G16B 20/00 (2019.01) G16H 50/20 (2018.01)

(52) CPC특허분류

G16H 50/70 (2018.01)

G01N 33/53 (2018.05)

(21) 출원번호 10-2019-0031623

(22) 출원일자 2019년03월20일

심사청구일자 2019년03월20일

(71) 출원인

인천대학교 산학협력단

인천광역시 연수구 아카데미로 119 (송도동)

(72) 발명자

안재균

서울특별시 서초구 서초대로74길 30, 우성아파트
501동 1314호

오일환

경기도 김포시 월곶면 애기봉로 3

(74) 대리인

김효성

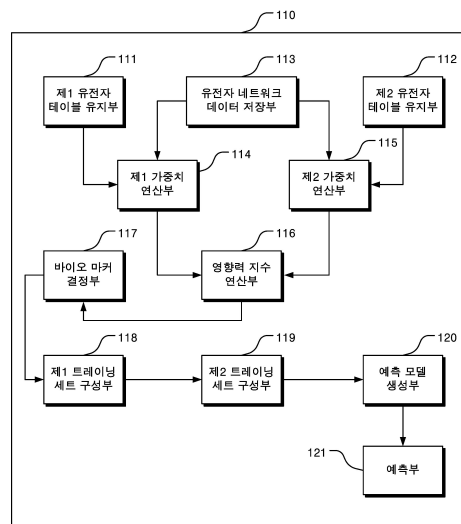
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 **생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치 및 그 동작 방법**

(57) 요약

생성적 적대 신경망(Generative Adversarial Networks: GANs)을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치 및 그 동작 방법이 개시된다. 본 발명은 암의 예후가 좋은 군과 나쁜 군으로 사전 분류된 암환자로부터 수집된 유전자들의 유전자 데이터의 종류별 데이터 값을 기초로 GANs를 이용한 기계학습을 통해 유전자 네트워크에서의 각 유전자 간의 영향력 정도를 나타내는 가중치를 산출하고, 산출된 가중치를 기초로 각 유전자의 암에 대한 영향력 지수를 산출한 후 영향력 지수가 높은 일부 유전자들을 암의 예후 예측을 위한 바이오 마커로 결정하고, 상기 결정된 바이오 마커를 이용하여 암의 예후 예측 모델을 만듦으로써, 환자의 유전자 특성에 기반하여 암의 예후를 예측하는 예측 모델의 예측 정확도를 높일 수 있다.

대표도 - 도1



(52) CPC특허분류

G16B 20/00 (2019.02)

G16H 50/20 (2018.01)

명세서

청구범위

청구항 1

암의 예후가 좋은 군으로 사전 분류되어 있는 암환자로부터 수집된 서로 다른 종류의 복수의 유전자들 각각에 대한 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제1 유전자 테이블을 저장하여 유지하는 제1 유전자 테이블 유지부;

암의 예후가 나쁜 군으로 사전 분류되어 있는 암환자로부터 수집된 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제2 유전자 테이블을 저장하여 유지하는 제2 유전자 테이블 유지부;

상기 복수의 유전자들 중 서로 영향을 미치는 유전자 간에 링크가 설정되어 있는 기설정된(predetermined) 유전자 네트워크에 대한 데이터가 저장되어 있는 유전자 네트워크 데이터 저장부;

상기 제1 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 생성적 적대 신경망(Generative Adversarial Networks: GANs)을 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제1 가중치를 연산하는 제1 가중치 연산부;

상기 제2 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 상기 GANs를 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제2 가중치를 연산하는 제2 가중치 연산부;

상기 복수의 유전자 데이터의 종류별로 상기 제1 가중치와 상기 제2 가중치가 연산되면, 상기 복수의 유전자 데이터의 종류별로, 상기 제1 가중치와 상기 제2 가중치에 기초하여 상기 복수의 유전자들 각각의 암에 대한 영향도를 나타내는 영향력 지수를 연산하는 영향력 지수 연산부; 및

상기 복수의 유전자들 중 상기 복수의 유전자 데이터의 종류별로 연산된 영향력 지수의 총 합계가 큰 순서로 기설정된 개수의 유전자들을 선정하고, 상기 선정된 유전자들을 암의 예후 예측을 위한 바이오 마커로 결정하는 바이오 마커 결정부

를 포함하는 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치.

청구항 2

제1항에 있어서,

상기 제1 가중치 연산부는

상기 기설정된 유전자 네트워크를 기초로 상기 GANs의 생성기(generator)를 통해 출력되는 상기 복수의 유전자 데이터의 종류별 모조(fake) 데이터 값이, 상기 GANs의 식별기(discriminator)를 통해 연산되는 상기 제1 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 일치할 가능성에 대한 확률이 최대가 되도록 상기 생성기와 상기 식별기에 대한 학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 상기 제1 가중치를 연산하고,

상기 제2 가중치 연산부는

상기 기설정된 유전자 네트워크를 기초로 상기 GANs의 생성기를 통해 출력되는 상기 복수의 유전자 데이터의 종류별 모조 데이터 값이, 상기 GANs의 식별기를 통해 연산되는 상기 제2 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 일치할 가능성에 대한 확률이 최대가 되도록 상기 생성기와 상기 식별기에 대한 학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로,

상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 상기 제2 가중치를 연산하는 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치.

청구항 3

제1항에 있어서,

상기 영향력 지수 연산부는

상기 복수의 유전자 데이터의 종류별로 상기 제1 가중치와 상기 제2 가중치가 연산되면, 상기 복수의 유전자 데이터의 종류별로, 상기 복수의 유전자들 각각에서의 상기 제1 가중치의 분포와 상기 제2 가중치의 분포 간의 쿨백-라이블러 발산(Kullback-Leibler Divergence)과 상기 제2 가중치의 분포와 상기 제1 가중치의 분포 간의 쿨백-라이블러 발산의 합계를 상기 복수의 유전자들 각각의 암에 대한 영향도를 나타내는 상기 영향력 지수로 연산하는 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치.

청구항 4

제1항에 있어서,

상기 기설정된 개수의 유전자들이 상기 바이오 마커로 결정된 이후, 사용자에게 의해 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각으로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록된 제1 바이오 마커 테이블과 암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각으로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록된 제2 바이오 마커 테이블이 입력되면서, 암의 예후 예측을 위한 모델 생성 명령이 인가되면, 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 상기 제1 바이오 마커 테이블에 기록된 데이터 값을 입력으로 지정하고, 암의 예후가 좋음을 지시하는 제1 출력 정보를 출력으로 지정함으로써, 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 제1 트레이닝 세트를 구성하는 제1 트레이닝 세트 구성부;

암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 상기 제2 바이오 마커 테이블에 기록된 데이터 값을 입력으로 지정하고, 암의 예후가 나쁨을 지시하는 제2 출력 정보를 출력으로 지정함으로써, 암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 제2 트레이닝 세트를 구성하는 제2 트레이닝 세트 구성부; 및

상기 제1 트레이닝 세트와 상기 제2 트레이닝 세트를 기초로 지도 학습(supervised learning) 기반의 기계학습을 수행함으로써, 암의 예후 예측 모델을 생성하는 예측 모델 생성부

를 더 포함하는 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치.

청구항 5

제4항에 있어서,

상기 암의 예후 예측 모델이 생성된 이후, 상기 사용자에게 의해 암의 예후 예측의 대상이 되는 제1 암환자로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 입력으로 인가되면서, 상기 제1 암환자에 대한 암의 예후 예측 명령이 인가되면, 상기 암의 예후 예측 모델에 상기 제1 암환자로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값을 입력으로 인가하여 상기 제1 암환자에 대한 암의 예후 예측 정보로 상기 제1 출력 정보 또는 상기 제2 출력 정보 중 어느 하나의 출력 정보를 산출하는 예측부

를 더 포함하는 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치.

청구항 6

암의 예후가 좋은 군으로 사전 분류되어 있는 암환자로부터 수집된 서로 다른 종류의 복수의 유전자들 각각에 대한 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제1 유전자 테이블을 저장하여 유지하는 단계;

암의 예후가 나쁜 군으로 사전 분류되어 있는 암환자로부터 수집된 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제2 유전자 테이블을 저장하여 유지하는 단계;

상기 복수의 유전자들 중 서로 영향을 미치는 유전자 간에 링크가 설정되어 있는 기설정된(predetermined) 유전자 네트워크에 대한 데이터가 저장되어 있는 유전자 네트워크 데이터 저장부를 유지하는 단계;

상기 제1 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 생성적 적대 신경망(Generative Adversarial Networks: GANs)을 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제1 가중치를 연산하는 단계;

상기 제2 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 상기 GANs를 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제2 가중치를 연산하는 단계;

상기 복수의 유전자 데이터의 종류별로 상기 제1 가중치와 상기 제2 가중치가 연산되면, 상기 복수의 유전자 데이터의 종류별로, 상기 제1 가중치와 상기 제2 가중치에 기초하여 상기 복수의 유전자들 각각의 암에 대한 영향도를 나타내는 영향력 지수를 연산하는 단계; 및

상기 복수의 유전자들 중 상기 복수의 유전자 데이터의 종류별로 연산된 영향력 지수의 총 합계가 큰 순서로 기설정된 개수의 유전자들을 선정하고, 상기 선정된 유전자들을 암의 예후 예측을 위한 바이오 마커로 결정하는 단계

를 포함하는 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법.

청구항 7

제6항에 있어서,

상기 제1 가중치를 연산하는 단계는

상기 기설정된 유전자 네트워크를 기초로 상기 GANs의 생성기(generator)를 통해 출력되는 상기 복수의 유전자 데이터의 종류별 모조(fake) 데이터 값이, 상기 GANs의 식별기(discriminator)를 통해 연산되는 상기 제1 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 일치할 가능성에 대한 확률이 최대가 되도록 상기 생성기와 상기 식별기에 대한 학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 상기 제1 가중치를 연산하고,

상기 제2 가중치를 연산하는 단계는

상기 기설정된 유전자 네트워크를 기초로 상기 GANs의 생성기를 통해 출력되는 상기 복수의 유전자 데이터의 종류별 모조 데이터 값이, 상기 GANs의 식별기를 통해 연산되는 상기 제2 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 일치할 가능성에 대한 확률이 최대가 되도록 상기 생성기와 상기 식별기에 대한 학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 상기 제2 가중치를 연산하는 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법.

청구항 8

제6항에 있어서,

상기 영향력 지수를 연산하는 단계는

상기 복수의 유전자 데이터의 종류별로 상기 제1 가중치와 상기 제2 가중치가 연산되면, 상기 복수의 유전자 데이터의 종류별로, 상기 복수의 유전자들 각각에서의 상기 제1 가중치의 분포와 상기 제2 가중치의 분포 간의 쿨

백-라이블러 발산(Kullback-Leibler Divergence)과 상기 제2 가중치의 분포와 상기 제1 가중치의 분포 간의 쿨백-라이블러 발산의 합계를 상기 복수의 유전자들 각각의 암에 대한 영향도를 나타내는 상기 영향력 지수로 연산하는 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법.

청구항 9

제6항에 있어서,

상기 기설정된 개수의 유전자들이 상기 바이오 마커로 결정된 이후, 사용자에게 의해 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각으로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록된 제1 바이오 마커 테이블과 암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각으로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록된 제2 바이오 마커 테이블이 입력되면서, 암의 예후 예측을 위한 모델 생성 명령이 인가되면, 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 상기 제1 바이오 마커 테이블에 기록된 데이터 값을 입력으로 지정하고, 암의 예후가 좋음을 지시하는 제1 출력 정보를 출력으로 지정함으로써, 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 제1 트레이닝 세트를 구성하는 단계;

암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 상기 제2 바이오 마커 테이블에 기록된 데이터 값을 입력으로 지정하고, 암의 예후가 나쁨을 지시하는 제2 출력 정보를 출력으로 지정함으로써, 암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 제2 트레이닝 세트를 구성하는 단계; 및

상기 제1 트레이닝 세트와 상기 제2 트레이닝 세트를 기초로 지도 학습(supervised learning) 기반의 기계학습을 수행함으로써, 암의 예후 예측 모델을 생성하는 단계

를 더 포함하는 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법.

청구항 10

제9항에 있어서,

상기 암의 예후 예측 모델이 생성된 이후, 상기 사용자에게 의해 암의 예후 예측의 대상이 되는 제1 암환자로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 입력으로 인가되면서, 상기 제1 암환자에 대한 암의 예후 예측 명령이 인가되면, 상기 암의 예후 예측 모델에 상기 제1 암환자로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값을 입력으로 인가하여 상기 제1 암환자에 대한 암의 예후 예측 정보로 상기 제1 출력 정보 또는 상기 제2 출력 정보 중 어느 하나의 출력 정보를 산출하는 단계

를 더 포함하는 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법.

청구항 11

제6항 내지 제10항 중 어느 한 항의 방법을 컴퓨터와의 결합을 통해 실행시키기 위한 컴퓨터 프로그램을 기록한 컴퓨터 판독 가능 기록 매체.

청구항 12

제6항 내지 제10항 중 어느 한 항의 방법을 컴퓨터와의 결합을 통해 실행시키기 위한 저장매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

본 발명은 생성적 적대 신경망(Generative Adversarial Networks: GANs)을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치 및 그 동작 방법에 대한 것이다.

[0001]

배경 기술

- [0002] 최근, 식생활의 서구화 등으로 인해 암환자가 증가함에 따라 암의 치료를 위한 다양한 방법이 강구되고 있다.
- [0003] 암을 치료하는데 있어서, 환자의 유전자별 특성에 따라 암의 예후를 미리 예측할 수 있다면, 환자에게 암의 예후에 따른 치료 방법을 적절히 적용함으로써, 암 치료의 효과를 극대화할 수 있을 것이다.
- [0004] 최근에는 인공지능 기술의 발전으로 인해 환자의 유전자별 특성에 따라 암의 예후를 예측할 수 있도록 하는 예측 모델의 도입도 고려되고 있다.
- [0005] 관련해서, 암환자들을 암의 예후가 좋은 군과 암의 예후가 나쁜 군으로 구분한 후 암의 예후가 좋은 군으로 분류된 암환자들의 유전자 특성과 암의 예후가 나쁜 군으로 분류된 암환자들의 유전자 특성을 기초로 기계학습을 수행함으로써, 특정 암환자의 유전자 특성을 입력으로 인가하였을 때, 해당 암환자의 암의 예후가 좋을 것인지 나쁠 것인지를 미리 예측할 수 있는 예측 모델의 구성을 고려할 수 있다.
- [0006] 하지만, 사람의 유전자는 그 종류가 너무 많기 때문에 모든 유전자 특성을 고려해서 암의 예후를 예측하는 예측 모델을 구성하는데에 한계가 존재한다. 아울러, 암의 예후에 영향을 크게 미치지 않는 유전자도 있기 때문에 모든 유전자 특성을 기초로 암의 예후를 예측하는 예측 모델을 구성하게 되면, 예측 모델의 정확도가 낮아질 수 있는 문제가 있다.
- [0007] 따라서, 많은 수의 유전자들 중에서 암의 예후에 영향을 미치는 특정 유전자들만을 바이오 마커로 선별하고, 선별된 바이오 마커를 기초로 암의 예후를 예측할 수 있는 예측 모델을 구성함으로써, 암의 예후 예측의 정확도를 높일 수 있는 기술의 연구가 필요하다.

발명의 내용

해결하려는 과제

- [0008] 본 발명에 따른 생성적 적대 신경망(Generative Adversarial Networks: GANs)을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치는 암의 예후가 좋은 군과 나쁜 군으로 사전 분류된 암환자로부터 수집된 유전자들의 유전자 데이터의 종류별 데이터 값을 기초로 GANs를 이용한 기계학습을 통해 유전자 네트워크에서의 각 유전자 간의 영향력 정도를 나타내는 가중치를 산출하고, 산출된 가중치를 기초로 각 유전자의 암에 대한 영향력 지수를 산출한 후 영향력 지수가 높은 일부 유전자들을 암의 예후 예측을 위한 바이오 마커로 결정하고, 상기 결정된 바이오 마커를 이용하여 암의 예후 예측 모델을 만듦으로써, 환자의 유전자 특성에 기반하여 암의 예후를 예측하는 예측 모델의 예측 정확도를 높일 수 있도록 한다.

과제의 해결 수단

- [0009] 본 발명의 일실시예에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치는 암의 예후가 좋은 군으로 사전 분류되어 있는 암환자로부터 수집된 서로 다른 종류의 복수의 유전자들 각각에 대한 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제1 유전자 테이블을 저장하여 유지하는 제1 유전자 테이블 유지부, 암의 예후가 나쁜 군으로 사전 분류되어 있는 암환자로부터 수집된 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제2 유전자 테이블을 저장하여 유지하는 제2 유전자 테이블 유지부, 상기 복수의 유전자들 중 서로 영향을 미치는 유전자 간에 링크가 설정되어 있는 기설정된(predetermined) 유전자 네트워크에 대한 데이터가 저장되어 있는 유전자 네트워크 데이터 저장부, 상기 제1 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 생성적 적대 신경망(Generative Adversarial Networks: GANs)을 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제1 가중치를 연산하는 제1 가중치 연산부, 상기 제2 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 상기 GANs를 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제2 가중치를 연산하는 제2 가중치 연산부, 상기 복수의 유전자 데이터의 종류별로 상기 제1 가중치와 상기 제2 가중치가 연산되면, 상기 복수의 유전자 데이터의 종류별로, 상기 제1 가중치와 상기 제2 가중치에 기초하여 상기 복수의 유전자들 각각의 암에 대한 영향도를 나타내는 영향력 지수를 연산하는 영향력 지수 연산부 및 상기 복수의 유

전자들 중 상기 복수의 유전자 데이터의 종류별로 연산된 영향력 지수의 총 합계가 큰 순서로 기설정된 개수의 유전자들을 선정하고, 상기 선정된 유전자들을 암의 예후 예측을 위한 바이오 마커로 결정하는 바이오 마커 결정부를 포함한다.

[0010] 또한, 본 발명의 일실시예에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법은 암의 예후가 좋은 군으로 사전 분류되어 있는 암환자로부터 수집된 서로 다른 종류의 복수의 유전자들 각각에 대한 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제1 유전자 테이블을 저장하여 유지하는 단계, 암의 예후가 나쁜 군으로 사전 분류되어 있는 암환자로부터 수집된 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제2 유전자 테이블을 저장하여 유지하는 단계, 상기 복수의 유전자들 중 서로 영향을 미치는 유전자 간에 링크가 설정되어 있는 기설정된 유전자 네트워크에 대한 데이터가 저장되어 있는 유전자 네트워크 데이터 저장부를 유지하는 단계, 상기 제1 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 GANs를 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제1 가중치를 연산하는 단계, 상기 제2 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 상기 GANs를 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제2 가중치를 연산하는 단계, 상기 복수의 유전자 데이터의 종류별로 상기 제1 가중치와 상기 제2 가중치가 연산되면, 상기 복수의 유전자 데이터의 종류별로, 상기 제1 가중치와 상기 제2 가중치에 기초하여 상기 복수의 유전자들 각각의 암에 대한 영향도를 나타내는 영향력 지수를 연산하는 단계 및 상기 복수의 유전자들 중 상기 복수의 유전자 데이터의 종류별로 연산된 영향력 지수의 총 합계가 큰 순서로 기설정된 개수의 유전자들을 선정하고, 상기 선정된 유전자들을 암의 예후 예측을 위한 바이오 마커로 결정하는 단계를 포함한다.

발명의 효과

[0011] 본 발명에 따른 생성적 적대 신경망(Generative Adversarial Networks: GANs)을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치는 암의 예후가 좋은 군과 나쁜 군으로 사전 분류된 암환자로부터 수집된 유전자들의 유전자 데이터의 종류별 데이터 값을 기초로 GANs를 이용한 기계학습을 통해 유전자 네트워크에서의 각 유전자 간의 영향력 정도를 나타내는 가중치를 산출하고, 산출된 가중치를 기초로 각 유전자의 암에 대한 영향력 지수를 산출한 후 영향력 지수가 높은 일부 유전자들을 암의 예후 예측을 위한 바이오 마커로 결정하고, 상기 결정된 바이오 마커를 이용하여 암의 예후 예측 모델을 만듦으로써, 환자의 유전자 특성에 기반하여 암의 예후를 예측하는 예측 모델의 예측 정확도를 높일 수 있다.

도면의 간단한 설명

[0012] 도 1은 본 발명의 일실시예에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 구조를 도시한 도면이다.

도 2와 도 3은 본 발명의 일실시예에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작을 설명하기 위한 도면이다.

도 4는 본 발명의 일실시예에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법을 도시한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0013] 이하에서는 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명하기로 한다. 이러한 설명은 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였으며, 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 본 명세서 상에서 사용되는 모든 용어들은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 사람에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다.

[0014] 본 문서에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있다는 것을 의미한다. 또한, 본 발명의 다양한

실시예들에 있어서, 각 구성요소들, 기능 블록들 또는 수단들은 하나 또는 그 이상의 하부 구성요소로 구성될 수 있고, 각 구성요소들이 수행하는 전기, 전자, 기계적 기능들은 전자회로, 집적회로, ASIC(Application Specific Integrated Circuit) 등 공지된 다양한 소자들 또는 기계적 요소들로 구현될 수 있으며, 각각 별개로 구현되거나 2 이상이 하나로 통합되어 구현될 수도 있다.

[0015] 한편, 첨부된 블록도의 블록들이나 흐름도의 단계들은 범용 컴퓨터, 특수용 컴퓨터, 휴대용 노트북 컴퓨터, 네트워크 컴퓨터 등 데이터 프로세싱이 가능한 장비의 프로세서나 메모리에 탑재되어 지정된 기능들을 수행하는 컴퓨터 프로그램 명령들(instructions)을 의미하는 것으로 해석될 수 있다. 이들 컴퓨터 프로그램 명령들은 컴퓨터 장치에 구비된 메모리 또는 컴퓨터에서 판독 가능한 메모리에 저장될 수 있기 때문에, 블록도의 블록들 또는 흐름도의 단계들에서 설명된 기능들은 이를 수행하는 명령 수단을 내포하는 제조물로 생산될 수도 있다. 아울러, 각 블록 또는 각 단계는 특정된 논리적 기능(들)을 실행하기 위한 하나 이상의 실행 가능한 명령들을 포함하는 모듈, 세그먼트 또는 코드의 일부를 나타낼 수 있다. 또, 몇 가지 대체 가능한 실시예들에서는 블록들 또는 단계들에서 언급된 기능들이 정해진 순서와 달리 실행되는 것도 가능함을 주목해야 한다. 예컨대, 잇달아 도시되어 있는 두 개의 블록들 또는 단계들은 실질적으로 동시에 수행되거나, 역순으로 수행될 수 있으며, 경우에 따라 일부 블록들 또는 단계들이 생략된 채로 수행될 수도 있다.

[0016] 도 1은 본 발명의 일실시예에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 구조를 도시한 도면이다.

[0017] 도 1을 참조하면, 본 발명에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치(110)는 제1 유전자 테이블 유지부(111), 제2 유전자 테이블 유지부(112), 유전자 네트워크 데이터 저장부(113), 제1 가중치 연산부(114), 제2 가중치 연산부(115), 영향력 지수 연산부(116) 및 바이오 마커 결정부(117)를 포함한다.

[0018] 제1 유전자 테이블 유지부(111)는 암의 예후가 좋은 군으로 사전 분류되어 있는 암환자로부터 수집된 서로 다른 종류의 복수의 유전자들 각각에 대한 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제1 유전자 테이블을 저장하여 유지한다.

[0019] 여기서, 서로 다른 복수의 유전자들이란 BRCA, LAML, LIHC, LUAD 등과 같은 유전자들을 의미하며, 유전자 데이터의 종류란 mRNA, CNV, Met, SNP 등과 같이 각 유전자에서 확인할 수 있는 데이터들의 의미한다.

[0020] 그리고, 암의 예후가 좋은 군이란 암환자들 중 사용자에게 의해 암의 예후가 좋은 것으로 판단되는 환자들만을 사전 분류한 군을 의미하는 것으로, 상기 제1 유전자 테이블에는 하기의 표 1과 같이 암의 예후가 좋은 군으로 사전 분류된 특정 암환자로부터 수집된 각 유전자에 대한 유전자 데이터의 종류별 데이터 값이 기록되어 있을 수 있다.

표 1

복수의 유전자들	복수의 유전자 데이터의 종류별 데이터 값			
	mRNA	CNV	Met	SNP
BRCA	11578	11093	9216	10345
LAML	11400	11234	12451	12421
LIHC	15451	12545	12355	15679
LUAD	11499	11634	11154	7577
...	...	...	...	...

[0024] 제2 유전자 테이블 유지부(112)는 암의 예후가 나쁜 군으로 사전 분류되어 있는 암환자로부터 수집된 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제2 유전자 테이블을 저장하여 유지한다.

[0025] 여기서, 암의 예후가 나쁜 군이란 암환자들 중 사용자에게 의해 암의 예후가 나쁜 것으로 판단되는 환자들만을 사전 분류한 군을 의미하는 것으로, 상기 제2 유전자 테이블에는 하기의 표 2와 같이 암의 예후가 나쁜 군으로 사

전 분류된 특정 암환자로부터 수집된 각 유전자에 대한 유전자 데이터의 종류별 데이터 값이 기록되어 있을 수 있다.

표 2

복수의 유전자들	복수의 유전자 데이터의 종류별 데이터 값			
	mRNA	CNV	Met	SNP
BRCA	12345	12356	12111	12577
LAML	15312	11234	12576	12457
LIHC	12345	12546	7890	12346
LUAD	12467	11134	12357	8907
...	...	...	...	...

유전자 네트워크 데이터 저장부(113)에는 상기 복수의 유전자들 중 서로 영향을 미치는 유전자 간에 링크가 설정되어 있는 기설정된(predetermined) 유전자 네트워크(genetic networks)에 대한 데이터가 저장되어 있다.

여기서, 유전자 네트워크란 도 2에 도시된 그림과 같이 서로 영향을 미치는 유전자 간에 링크가 설정된 정보를 의미하는 것으로, 이러한 유전자 네트워크는 생물학적 경로, protein-protein interaction(PPI), Gene Ontology(GO) 데이터 등에 의해서 구축될 수 있다.

제1 가중치 연산부(114)는 상기 제1 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 생성적 적대 신경망(Generative Adversarial Networks: GANs)을 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제1 가중치를 연산한다.

제2 가중치 연산부(115)는 상기 제2 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 상기 GANs를 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제2 가중치를 연산한다.

여기서, GANs란 비지도 학습에 사용되는 인공지능 알고리즘으로, 생성기(generator)와 식별기(discriminator)로 구성되어 있어, 생성기가 모조(fake) 데이터를 생성하면, 식별기에서 실제(real) 데이터를 기초로 상기 모조 데이터가 실제인지 모조인지 여부를 확률적으로 검토하는 과정을 반복 학습함으로써, 결국 생성기가 실제 데이터와 거의 유사한 모조 데이터를 생성하도록 구성하는 알고리즘을 의미한다.

관련해서, 본 발명의 일실시예에 따르면, 제1 가중치 연산부(114)는 상기 기설정된 유전자 네트워크를 기초로 상기 GANs의 생성기를 통해 출력되는 상기 복수의 유전자 데이터의 종류별 모조 데이터 값, 상기 GANs의 식별기를 통해 연산되는 상기 제1 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 일치할 가능성에 대한 확률이 최대가 되도록 상기 생성기와 상기 식별기에 대한 학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 상기 제1 가중치를 연산할 수 있다.

그리고, 제2 가중치 연산부(115)는 상기 기설정된 유전자 네트워크를 기초로 상기 GANs의 생성기를 통해 출력되는 상기 복수의 유전자 데이터의 종류별 모조 데이터 값, 상기 GANs의 식별기를 통해 연산되는 상기 제2 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 일치할 가능성에 대한 확률이 최대가 되도록 상기 생성기와 상기 식별기에 대한 학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 상기 제2 가중치를 연산할 수 있다.

이하에서는 도 3을 참조하여 제1 가중치 연산부(114)가 상기 제1 가중치를 연산하는 과정에 대해 예를 들어 상세히 설명하기로 한다.

[0037] 우선, 암의 예후가 좋은 군으로 사전 분류되어 있는 암환자로부터 수집된 유전자로 'a, b, c' 3개가 존재한다고 가정하고, 상기 유전자 'a, b, c' 각각의 유전자 데이터의 종류 중 mRNA에 대한 데이터 값을 기초로 mRNA에 대한 제1 가중치를 연산하는 상황을 가정하며, 상기 유전자 'a, b, c' 각각의 mRNA에 대한 데이터 값이 '0.4, 0.3, 0.7'이라고 가정하자.

[0038] 그리고, 상기 유전자 'a, b, c'에 대한 유전자 네트워크가 도면부호 300으로 표시한 그림과 같다고 가정하자.

[0039] 우선, 유전자 네트워크(300)는 도면부호 330으로 표시한 그림과 같은 인접 행렬로 나타낼 수 있다. 인접 행렬(330)은 행과 열의 항목을 모두 'a, b, c'라고 하였을 때, 유전자 네트워크 상에서 서로 링크가 설정된 유전자 간의 성분을 '1'로, 링크가 설정되지 않은 유전자 간의 성분을 '0'으로 채워서 3x3 크기의 행렬로 구성할 수 있다.

[0040] 이때, 제1 가중치 연산부(114)는 인접 행렬(330)을 'A'라고 하고, 상기 제1 가중치를 나타내는 행렬을 'W'라고 하였을 때, 생성기(311)를 이용해서 도면부호 321로 나타낸 그림과 같이, $A \odot W \odot W^T$ 를 연산할 수 있다.

[0041] 여기서, $\odot$ 는 아다마르 곱(Hadamard product)을 의미하고, W^T 는 상기 제1 가중치를 나타내는 행렬 W의 전치 행렬을 의미한다.

[0042] 그리고 나서, 제1 가중치 연산부(114)는 도면부호 310에 나타낸 바와 같이, 평균이 '0'이고, 분산이 '1'인 표준 정규분포에서 3개의 임의의 값을 추출하고, 3개의 임의의 값을 성분으로 포함하는 3차원 벡터 X를 구성한 후 생성기(311)를 이용하여 도면부호 322에 나타낸 그림과 같이, 상기 3차원 벡터 X와 $A \odot W \odot W^T$ 간의 곱인 $X(A \odot W \odot W^T)$ 를 연산할 수 있다.

[0043] 그 이후, 제1 가중치 연산부(114)는 생성기(311)를 이용하여 $X(A \odot W \odot W^T)$ 에 대해 도면부호 323에 나타낸 그림과 같이, 하기의 수학적 식 1에 따른 연산을 수행함으로써, 상기 유전자 'a, b, c' 각각의 mRNA에 대한 모조 데이터 값을 출력할 수 있다.

수학적 식 1

$$\sigma(X(A \odot W \odot W^T)) = \frac{1}{1 + e^{-X(A \odot W \odot W^T)}}$$

[0047] 도 3에 도시된 그림과 같이 생성기(311)를 통해 '0.526, 0.502, 0.534'라는 모조 데이터 값이 출력되었다고 하는 경우, 제1 가중치 연산부(114)는 식별기(312)를 이용하여 '0.526, 0.502, 0.534'라는 모조 데이터 값이 상기 유전자 'a, b, c' 각각의 mRNA에 대한 실제 데이터 값인 '0.4, 0.3, 0.7'과의 일치여부에 대한 확률을 출력할 수 있다.

[0048] 이때, 제1 가중치 연산부(114)는 식별기(312)를 통해 출력되는 확률이 최대가 될 때까지 생성기(311)를 통한 모조 데이터 값의 출력 및 식별기(312)를 통한 확률 검증 과정을 반복 수행 학습함으로써, 생성기(311)에서 출력되는 모조 데이터 값이 상기 유전자 'a, b, c' 각각의 mRNA에 대한 실제 데이터 값인 '0.4, 0.3, 0.7'과 거의 일치하게 되도록 상기 제1 가중치를 나타내는 행렬 W를 찾아낼 수 있다.

[0049] 이렇게, 상기 제1 가중치를 나타내는 행렬 W를 찾아내면 제1 가중치 연산부(114)는 유전자 네트워크(300)를 참조하여 상기 행렬 W에 포함된 성분들 중 서로 영향을 미치는 유전자 간에서 매칭되는 성분을 유전자 네트워크(300) 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 상기 제1 가중치로 최종 확정

할 수 있다.

- [0050] 예컨대, 'a'라는 유전자는 유전자 네트워크(300)를 참조하면 'b' 및 'c'와 영향을 주고받는 유전자이기 때문에 제1 가중치 연산부(114)는 상기 행렬 W 에 포함된 성분들 중 'a'와 'b' 간에 매칭된 성분을 mRNA라는 유전자 데이터의 종류에서의 'a'의 'b'에 대한 상기 제1 가중치로 확정할 수 있고, 'a'와 'c' 간에 매칭된 성분을 mRNA라는 유전자 데이터의 종류에서의 'a'의 'c'에 대한 상기 제1 가중치로 확정할 수 있다.
- [0051] 그리고, 'b'라는 유전자는 유전자 네트워크(300)를 참조하면 'a'와 영향을 주고받는 유전자이기 때문에 제1 가중치 연산부(114)는 상기 행렬 W 에 포함된 성분들 중 'b'와 'a' 간에 매칭된 성분을 mRNA라는 유전자 데이터의 종류에서의 'b'의 'a'에 대한 상기 제1 가중치로 확정할 수 있다.
- [0052] 마지막으로, 'c'라는 유전자는 유전자 네트워크(300)를 참조하면 'a'와 영향을 주고받는 유전자이기 때문에 제1 가중치 연산부(114)는 상기 행렬 W 에 포함된 성분들 중 'c'와 'a' 간에 매칭된 성분을 mRNA라는 유전자 데이터의 종류에서의 'c'의 'a'에 대한 상기 제1 가중치로 확정할 수 있다.
- [0053] 이러한 방식으로, 제1 가중치 연산부(114)는 상기 유전자 'a, b, c'에 대해 mRNA라는 유전자 데이터의 종류뿐 아니라, CNV, Met, SNP 등의 다른 유전자 데이터의 종류에 대해서도 각 데이터 값에 기초하여 상기 제1 가중치를 연산할 수 있다.
- [0054] 그리고, 제2 가중치 연산부(115)도 제1 가중치 연산부(114)의 동작과 동일한 방식으로 암의 예후가 나쁜 군으로 사전 분류되어 있는 암환자로부터 수집된 유전자 각각의 유전자 데이터의 종류별 데이터 값이 기록된 제2 유전자 테이블과 유전자 네트워크에 기초하여 mRNA, CNV, Met, SNP 등의 유전자 데이터의 종류별로, 상기 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 묘사하는 상기 제2 가중치를 연산할 수 있다.
- [0055] 이렇게, 상기 복수의 유전자 데이터의 종류별로 상기 제1 가중치와 상기 제2 가중치가 연산되면, 영향력 지수 연산부(116)는 상기 복수의 유전자 데이터의 종류별로, 상기 제1 가중치와 상기 제2 가중치에 기초하여 상기 복수의 유전자들 각각의 암에 대한 영향도를 나타내는 영향력 지수를 연산한다.
- [0056] 이때, 본 발명의 일실시예에 따르면, 영향력 지수 연산부(116)는 상기 복수의 유전자 데이터의 종류별로 상기 제1 가중치와 상기 제2 가중치가 연산되면, 상기 복수의 유전자 데이터의 종류별로, 상기 복수의 유전자들 각각에서의 상기 제1 가중치의 분포와 상기 제2 가중치의 분포 간의 쿨백-라이블러 발산(Kullback-Leibler Divergence)과 상기 제2 가중치의 분포와 상기 제1 가중치의 분포 간의 쿨백-라이블러 발산의 합계를 상기 복수의 유전자들 각각의 암에 대한 영향도를 나타내는 상기 영향력 지수로 연산할 수 있다.
- [0057] 관련해서, 영향력 지수 연산부(116)는 하기의 수학적 2에 기초하여 상기 복수의 유전자 데이터의 종류별로, 상기 복수의 유전자들 각각의 암에 대한 영향도를 나타내는 상기 영향력 지수를 연산할 수 있다.

수학적 2

$$score = D_{KL}(P(x) \parallel Q(x)) + D_{KL}(Q(x) \parallel P(x))$$

[0059]

[0061] 여기서, score는 상기 영향력 지수, $P(x)$ 는 상기 제1 가중치의 분포, $Q(x)$ 는 상기 제2 가중치의 분포, $D_{KL}(P(x) \parallel Q(x))$ 은 상기 제1 가중치의 분포와 상기 제2 가중치의 분포 간의 쿨백-라이블러 발산, $D_{KL}(Q(x) \parallel P(x))$ 은 상기 제2 가중치의 분포와 상기 제1 가중치의 분포 간의 쿨백-라이블러 발산을 의미한다.

- [0062] 예컨대, 전술한 예에서와 같이, 유전자 'a, b, c'가 있고, mRNA라는 유전자 데이터의 종류에 대해서 'a'의 'b'에 대한 제1 가중치 및 'a'의 'c'에 대한 제1 가중치와 'a'의 'b'에 대한 제2 가중치 및 'a'의 'c'에 대한 제2 가중치가 연산되었다고 하는 경우, 영향력 지수 연산부(116)는 'a'의 'b'에 대한 제1 가중치 및 'a'의 'c'에 대

한 제1 가중치의 분포와 'a'의 'b'에 대한 제2 가중치 및 'a'의 'c'에 대한 제2 가중치의 분포를 이용하여 $D_{KL}(P(x) \parallel Q(x))$ 와 $D_{KL}(Q(x) \parallel P(x))$ 를 각각 연산한 후 연산된 값의 합계를 mRNA 라는 유전자 데이터의 종류에서의 유전자 'a'의 영향력 지수로 연산할 수 있다.

[0063] 결국, 영향력 지수 연산부(116)를 통해 연산되는 상기 복수의 유전자 데이터의 종류별 각 유전자의 영향력 지수는 하기의 표 3과 같이 나타낼 수 있다.

표 3

복수의 유전자들	복수의 유전자 데이터의 종류별 영향력 지수			
	mRNA	CNV	Met	SNP
BRCA	지수 1	지수 2	지수 3	지수 4
LAML	지수 5	지수 6	지수 7	지수 8
LIHC	지수 9	지수 10	지수 11	지수 12
LUAD	지수 13	지수 14	지수 15	지수 16
...	...	...	...	...

[0067] 바이오 마커 결정부(117)는 상기 복수의 유전자들 중 상기 복수의 유전자 데이터의 종류별로 연산된 영향력 지수의 총 합계가 큰 순서로 기설정된 개수의 유전자들을 선정하고, 상기 선정된 유전자들을 암의 예후 예측을 위한 바이오 마커로 결정한다.

[0068] 예컨대, 상기 기설정된 개수가 '5개'이고, 상기 복수의 유전자 데이터의 종류별 각 유전자에 대한 영향력 지수가 표 3과 같이 연산되었다고 하는 경우, 바이오 마커 결정부(117)는 상기 복수의 유전자들 중 상기 복수의 유전자 데이터의 종류별로 연산된 영향력 지수의 총 합계가 큰 순서로 '5개'의 유전자들을 선정하고, 상기 선정된 '5개'의 유전자들은 암에 영향을 많이 미치는 유전자라고 볼 수 있기 때문에 바이오 마커 결정부(117)는 상기 '5개'의 유전자들을 암의 예후 예측을 위한 바이오 마커로 결정할 수 있다.

[0069] 본 발명의 일실시예에 따르면, 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치(110)는 상기 바이오 마커의 결정이 완료되면, 상기 바이오 마커를 이용해서 암의 예후 예측을 위한 예측 모델을 만들기 위한 구성으로, 제1 트레이닝 세트 구성부(118), 제2 트레이닝 세트 구성부(119) 및 예측 모델 생성부(120)를 더 포함할 수 있다.

[0070] 제1 트레이닝 세트 구성부(118)는 상기 기설정된 개수의 유전자들이 상기 바이오 마커로 결정된 이후, 사용자에게 의해 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각으로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록된 제1 바이오 마커 테이블과 암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각으로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록된 제2 바이오 마커 테이블이 입력되면서, 암의 예후 예측을 위한 모델 생성 명령이 인가되면, 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 상기 제1 바이오 마커 테이블에 기록된 데이터 값을 입력으로 지정하고, 암의 예후가 나쁨을 지시하는 제1 출력 정보를 출력으로 지정함으로써, 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 제1 트레이닝 세트를 구성한다.

[0071] 즉, 제1 트레이닝 세트 구성부(118)는 암의 예후가 좋은 군의 환자들로부터 수집한 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값을 입력으로 지정하고, 암의 예후가 나쁨을 지시하는 제1 출력 정보를 출력으로 지정하여 상기 제1 트레이닝 세트를 구성할 수 있다.

[0072] 제2 트레이닝 세트 구성부(118)는 암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 상기 제2 바이오 마커 테이블에 기록된 데이터 값을 입력으로 지정하고, 암의 예후가 나쁨을 지시하는 제2 출력 정보를 출력으로 지정함으로써, 암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 제2 트레이닝 세트를 구성한다.

- [0073] 즉, 제2 트레이닝 세트 구성부(119)는 암의 예후가 나쁜 군의 환자들로부터 수집한 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값을 입력으로 지정하고, 암의 예후가 나쁨을 지시하는 제2 출력 정보를 출력으로 지정하여 상기 제2 트레이닝 세트를 구성할 수 있다.
- [0074] 이렇게, 상기 제1 트레이닝 세트와 상기 제2 트레이닝 세트가 구성되면, 예측 모델 생성부(120)는 상기 제1 트레이닝 세트와 상기 제2 트레이닝 세트를 기초로 지도 학습(supervised learning) 기반의 기계학습을 수행함으로써, 암의 예후 예측 모델을 생성할 수 있다.
- [0075] 여기서, 기계학습에 사용되는 알고리즘은 다양한 신경망 모델이 사용될 수 있다.
- [0076] 이때, 본 발명의 일실시예에 따르면, 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치(110)는 예측부(121)를 더 포함할 수 있다.
- [0077] 예측부(121)는 상기 암의 예후 예측 모델이 생성된 이후, 상기 사용자에게 의해 암의 예후 예측의 대상이 되는 제1 암환자로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 입력으로 인가되면서, 상기 제1 암환자에 대한 암의 예후 예측 명령이 인가되면, 상기 암의 예후 예측 모델에 상기 제1 암환자로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값을 입력으로 인가하여 상기 제1 암환자에 대한 암의 예후 예측 정보로 상기 제1 출력 정보 또는 상기 제2 출력 정보 중 어느 하나의 출력 정보를 산출한다.
- [0078] 결국, 본 발명에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치(110)는 암의 예후가 좋은 군과 나쁜 군으로 사전 분류된 암환자로부터 수집된 유전자들의 유전자 데이터의 종류별 데이터 값을 기초로 GANs를 이용한 기계학습을 통해 유전자 네트워크에서의 각 유전자 간의 영향력 정도를 나타내는 가중치를 산출하고, 산출된 가중치를 기초로 각 유전자의 암에 대한 영향력 지수를 산출한 후 영향력 지수가 높은 일부 유전자들을 암의 예후 예측을 위한 바이오 마커로 결정하고, 상기 결정된 바이오 마커를 이용하여 암의 예후 예측 모델을 만듦으로써, 환자의 유전자 특성에 기반하여 암의 예후를 예측하는 예측 모델의 예측 정확도를 높일 수 있다.
- [0079] 도 4는 본 발명의 일실시예에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법을 도시한 순서도이다.
- [0080] 단계(S410)에서는 암의 예후가 좋은 군으로 사전 분류되어 있는 암환자로부터 수집된 서로 다른 종류의 복수의 유전자들 각각에 대한 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제1 유전자 테이블을 저장하여 유지한다.
- [0081] 단계(S420)에서는 암의 예후가 나쁜 군으로 사전 분류되어 있는 암환자로부터 수집된 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록되어 있는 제2 유전자 테이블을 저장하여 유지한다.
- [0082] 단계(S430)에서는 상기 복수의 유전자들 중 서로 영향을 미치는 유전자 간에 링크가 설정되어 있는 기설정된 유전자 네트워크에 대한 데이터가 저장되어 있는 유전자 네트워크 데이터 저장부를 유지한다.
- [0083] 단계(S440)에서는 상기 제1 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 GANs을 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제1 가중치를 연산한다.
- [0084] 단계(S450)에서는 상기 제2 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 상기 기설정된 유전자 네트워크에 대한 데이터를 기초로 상기 GANs를 통한 기계학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 제2 가중치를 연산한다.
- [0085] 단계(S460)에서는 상기 복수의 유전자 데이터의 종류별로 상기 제1 가중치와 상기 제2 가중치가 연산되면, 상기 복수의 유전자 데이터의 종류별로, 상기 제1 가중치와 상기 제2 가중치에 기초하여 상기 복수의 유전자들 각각의 암에 대한 영향도를 나타내는 영향력 지수를 연산한다.
- [0086] 단계(S470)에서는 상기 복수의 유전자들 중 상기 복수의 유전자 데이터의 종류별로 연산된 영향력 지수의 총 합계가 큰 순서로 기설정된 개수의 유전자들을 선정하고, 상기 선정된 유전자들을 암의 예후 예측을 위한 바이오

마커로 결정한다.

- [0087] 이때, 본 발명의 일실시예에 따르면, 단계(S440)에서는 상기 기설정된 유전자 네트워크를 기초로 상기 GANs의 생성기를 통해 출력되는 상기 복수의 유전자 데이터의 종류별 모조 데이터 값이, 상기 GANs의 식별기를 통해 연산되는 상기 제1 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 일치할 가능성에 대한 확률이 최대가 되도록 상기 생성기와 상기 식별기에 대한 학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 상기 제1 가중치를 연산할 수 있다.
- [0088] 이때, 단계(S450)에서는 상기 기설정된 유전자 네트워크를 기초로 상기 GANs의 생성기를 통해 출력되는 상기 복수의 유전자 데이터의 종류별 모조 데이터 값이, 상기 GANs의 식별기를 통해 연산되는 상기 제2 유전자 테이블에 기록되어 있는 상기 복수의 유전자들 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값과 일치할 가능성에 대한 확률이 최대가 되도록 상기 생성기와 상기 식별기에 대한 학습을 수행함으로써, 상기 복수의 유전자 데이터의 종류별로, 상기 기설정된 유전자 네트워크 상에서 링크가 설정되어 있는 유전자 간의 영향을 미치는 정도를 표상하는 상기 제2 가중치를 연산할 수 있다.
- [0089] 또한, 본 발명의 일실시예에 따르면, 단계(S460)에서는 상기 복수의 유전자 데이터의 종류별로 상기 제1 가중치와 상기 제2 가중치가 연산되면, 상기 복수의 유전자 데이터의 종류별로, 상기 복수의 유전자들 각각에서의 상기 제1 가중치의 분포와 상기 제2 가중치의 분포 간의 쿨백-라이블러 발산과 상기 제2 가중치의 분포와 상기 제1 가중치의 분포 간의 쿨백-라이블러 발산의 합계를 상기 복수의 유전자들 각각의 암에 대한 영향도를 나타내는 상기 영향력 지수로 연산할 수 있다.
- [0090] 또한, 본 발명의 일실시예에 따르면, 상기 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법은 상기 기설정된 개수의 유전자들이 상기 바이오 마커로 결정된 이후, 사용자에게 의해 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각으로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록된 제1 바이오 마커 테이블과 암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각으로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 기록된 제2 바이오 마커 테이블이 입력되면서, 암의 예후 예측을 위한 모델 생성 명령이 인가되면, 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 상기 제1 바이오 마커 테이블에 기록된 데이터 값을 입력으로 지정하고, 암의 예후가 좋음을 지시하는 제1 출력 정보를 출력으로 지정함으로써, 암의 예후가 좋은 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 제1 트레이닝 세트를 구성하는 단계, 암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 상기 제2 바이오 마커 테이블에 기록된 데이터 값을 입력으로 지정하고, 암의 예후가 나쁨을 지시하는 제2 출력 정보를 출력으로 지정함으로써, 암의 예후가 나쁜 군으로 사전 분류되어 있는 복수의 암환자들 각각에 대한 제2 트레이닝 세트를 구성하는 단계 및 상기 제1 트레이닝 세트와 상기 제2 트레이닝 세트를 기초로 지도 학습 기반의 기계학습을 수행함으로써, 암의 예후 예측 모델을 생성하는 단계를 더 포함할 수 있다.
- [0091] 이때, 본 발명의 일실시예에 따르면, 상기 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법은 상기 암의 예후 예측 모델이 생성된 이후, 상기 사용자에게 의해 암의 예후 예측의 대상이 되는 제1 암환자로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값이 입력으로 인가되면서, 상기 제1 암환자에 대한 암의 예후 예측 명령이 인가되면, 상기 암의 예후 예측 모델에 상기 제1 암환자로부터 수집된 상기 바이오 마커 각각에 대한 상기 복수의 유전자 데이터의 종류별 데이터 값을 입력으로 인가하여 상기 제1 암환자에 대한 암의 예후 예측 정보로 상기 제1 출력 정보 또는 상기 제2 출력 정보 중 어느 하나의 출력 정보를 산출하는 단계를 더 포함할 수 있다.
- [0092] 이상, 도 4를 참조하여 본 발명의 일실시예에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법에 대해 설명하였다. 여기서, 본 발명의 일실시예에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법은 도 1을 이용하여 설명한 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치(110)의 동작에 대한 구성과 대응될 수 있으므로, 이에 대한 보다 상세한 설명은 생략하기로 한다.
- [0093] 본 발명의 일실시예에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법은 컴퓨터와의 결합을 통해 실행시키기 위한 저장매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다.

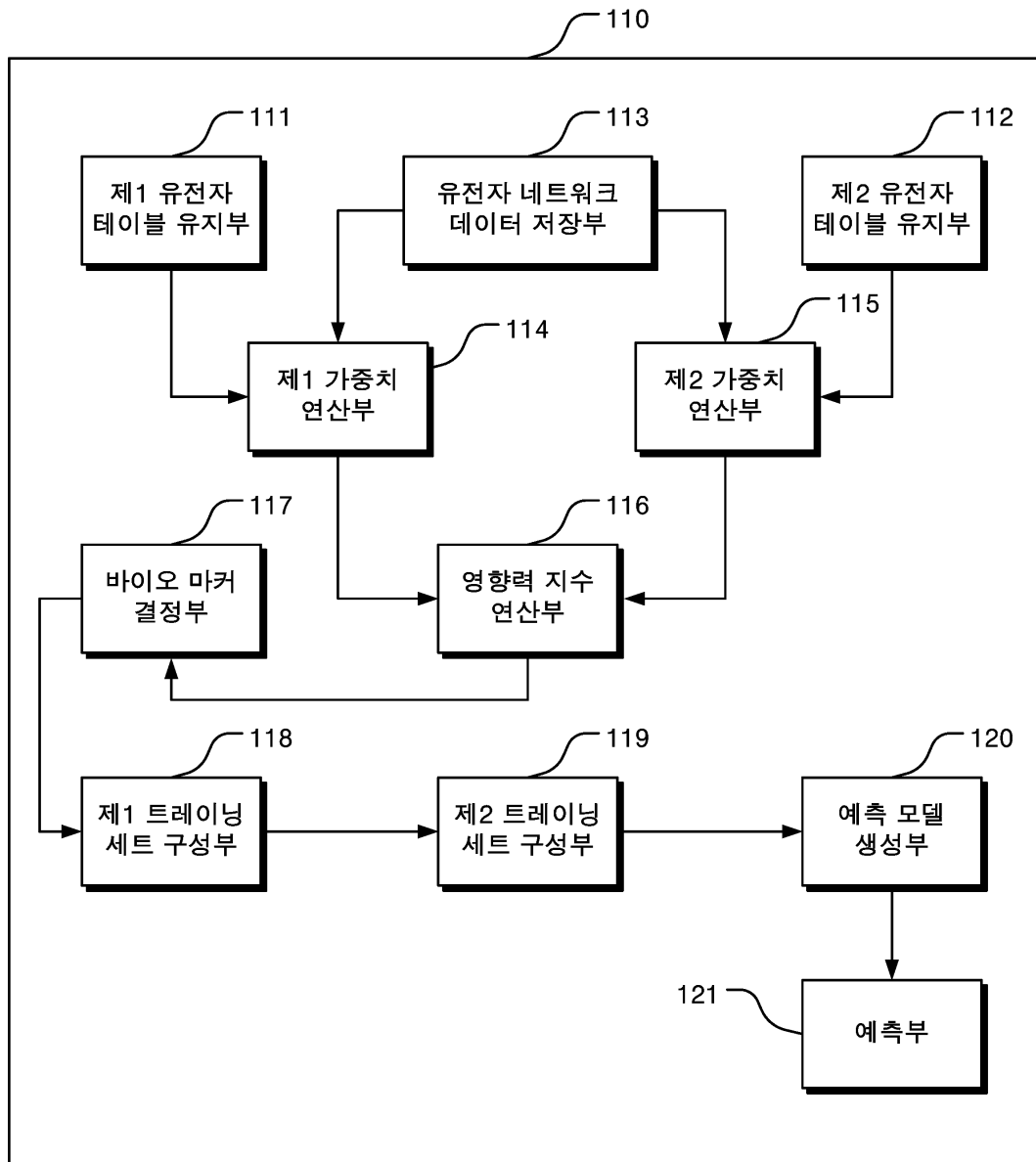
- [0094] 또한, 본 발명의 일실시에에 따른 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치의 동작 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 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.
- [0095] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.
- [0096] 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐만 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

부호의 설명

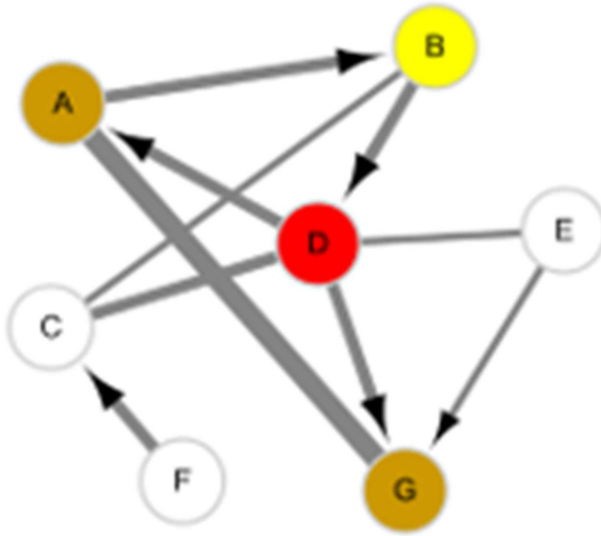
- [0097] 110: 생성적 적대 신경망을 기초로 암의 예후 예측에 사용되는 바이오 마커의 선정이 가능한 전자 장치
 111: 제1 유전자 테이블 유지부 112: 제2 유전자 테이블 유지부
 113: 유전자 네트워크 데이터 저장부 114: 제1 가중치 연산부
 115: 제2 가중치 연산부 116: 영향력 지수 연산부
 117: 바이오 마커 결정부 118: 제1 트레이닝 세트 구성부
 119: 제2 트레이닝 세트 구성부 120: 예측 모델 생성부
 121: 예측부

도면

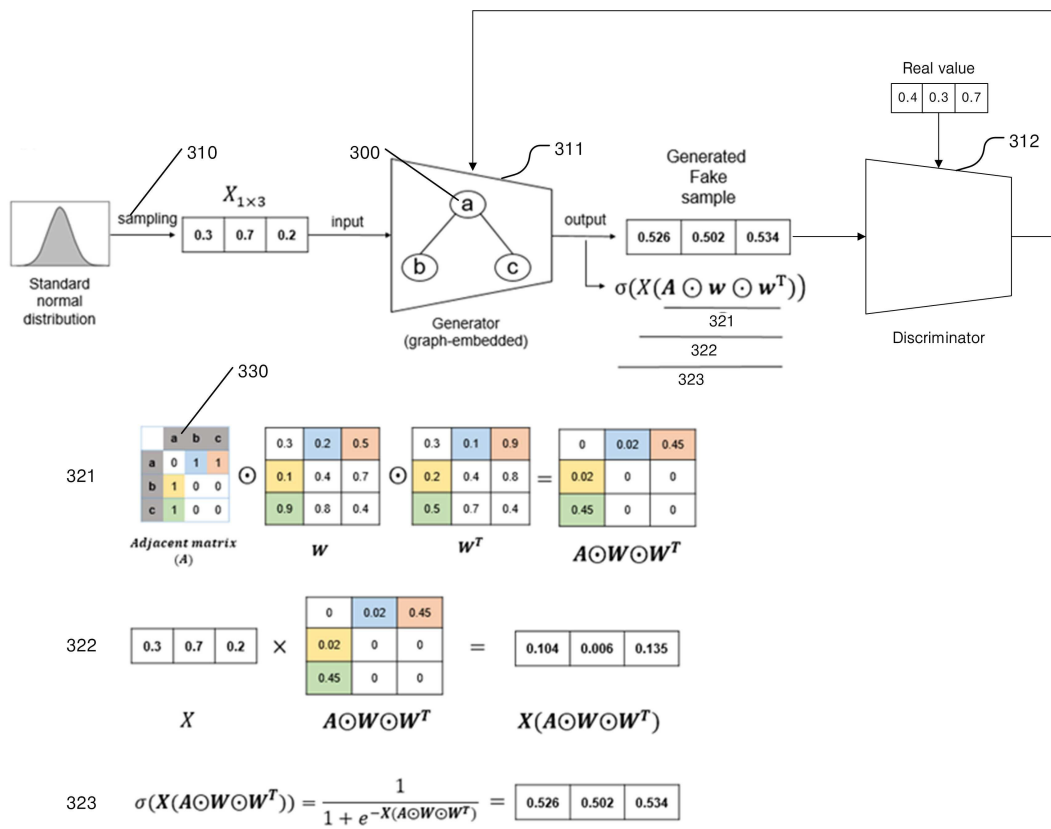
도면1



도면2



도면3



도면4

