



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2024년09월19일

(11) 등록번호 10-2706355

(24) 등록일자 2024년09월09일

(51) 국제특허분류(Int. Cl.)
G16H 50/20 (2018.01) A61B 5/00 (2021.01)
C12Q 1/6886 (2018.01) G06N 3/08 (2023.01)
G16B 20/00 (2019.01) G16B 25/10 (2019.01)
G16B 40/00 (2019.01)

(52) CPC특허분류
G16H 50/20 (2018.01)
A61B 5/7264 (2023.08)

(21) 출원번호 10-2021-0170287

(22) 출원일자 2021년12월01일

심사청구일자 2021년12월01일

(65) 공개번호 10-2023-0082734

(43) 공개일자 2023년06월09일

(56) 선행기술조사문헌

KR1020180036622 A*

KR1020210059325 A*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

부산대학교 산학협력단

부산광역시 금정구 부산대학로63번길 2 (장전동, 부산대학교)

(72) 발명자

황재준

경상남도 양산시 물금읍 금오로 20 양산부산대학교병원

김윤학

부산광역시 동래구 아시아드대로134번길

정성도

대구광역시 북구 동북로37길 10

(74) 대리인

오위환, 정기택

전체 청구항 수 : 총 14 항

심사관 : 이재균

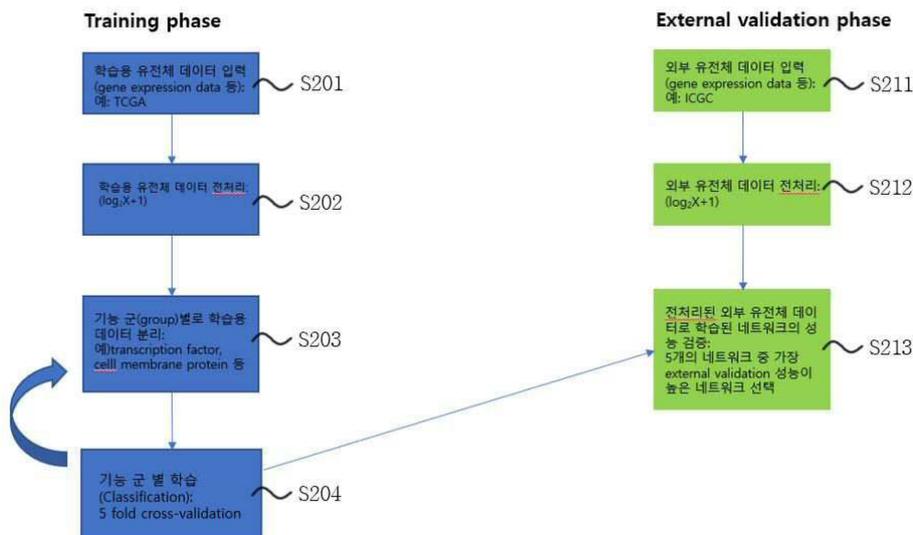
(54) 발명의 명칭 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법

(57) 요약

본 발명은 특정 기능을 하는 군(group)을 개별적으로 딥러닝 학습하여 그 중 가장 분류 정확도가 높은 몇 개의 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 한 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법에 관한 것으로, 학습용 유전체 데이터로 사용하기 위한 유전자 발현 데이터(gene

(뒷면에 계속)

대표도 - 도2



expression data)를 입력받아 학습용 유전체 데이터를 전처리하고 교차 검증을 위한 기능 군 별 학습 (Classification)을 수행하는 유전체 데이터 학습부;외부 유전체 데이터로 사용하기 위한 유전체 발현 데이터 (gene expression data)를 입력받아 외부 유전체 데이터를 전처리하고, 전처리된 외부 유전체 데이터로 학습된 네트워크의 성능을 검증하여 외부 유효성(external validation) 성능이 가장 높은 네트워크를 선택하는 외부 유전체 데이터 유효성 검사부;를 포함하고, 특정 기능을 하는 유전체 데이터 군(group)을 개별적으로 딥러닝 학습 하여 선택된 유전체 데이터 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 하는 것이다.

(52) CPC특허분류

- C12Q 1/6886* (2022.01)
- G06N 3/08* (2023.01)
- G16B 20/00* (2019.02)
- G16B 25/10* (2019.02)
- G16B 40/00* (2019.02)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711133925
과제번호	2018R1A5A2023879
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	집단연구지원(R&D)
연구과제명	치주질환신호 네트워크 연구센터
기 여 율	70/100
과제수행기관명	부산대학교
연구기간	2021.03.01 ~ 2022.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1711107213
과제번호	2020R1C1C1003741
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	GLUT2 표적 간세포암 조영제 개발 및 간세포암에서의 GLUT2 역할 규명
기 여 율	30/100
과제수행기관명	부산대학교
연구기간	2020.03.01 ~ 2021.02.28

명세서

청구범위

청구항 1

학습용 유전체 데이터로 사용하기 위한 유전자 발현 데이터(gene expression data)를 입력받아 학습용 유전체 데이터를 전처리하고 교차 검증을 위한 기능 군 별 학습(Classification)을 수행하는 유전체 데이터 학습부;

외부 유전체 데이터로 사용하기 위한 유전체 발현 데이터(gene expression data)를 입력받아 외부 유전체 데이터를 전처리하고, 전처리된 외부 유전체 데이터로 학습된 네트워크의 성능을 검증하여 외부 유효성(external validation) 성능이 가장 높은 네트워크를 선택하는 외부 유전체 데이터 유효성 검사부;를 포함하고,

유전체 발현 데이터(gene expression data)의 TF(transcription factor) 또는 세포막 유전체(cell membrane gene) 또는 microRNA의 특정 기능을 하는 군(group)을 개별적으로 딥러닝 학습하여 그 중 가장 분류 정확도 높은 군을 선택하여 딥러닝을 이용한 암종 분류(cancer type classification)의 대표 데이터로 사용하는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치.

청구항 2

제 1 항에 있어서, 유전체 데이터 학습부에 입력되는 유전체 발현 데이터(gene expression data)는 TCGA(The Cancer Genome Atlas)이거나,

외부 유전체 데이터 유효성 검사부에 입력되는 유전체 발현 데이터(gene expression data)는 ICGC(Internal Cancer Genome Consortium)인 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치.

청구항 3

제 1 항에 있어서, 유전체 데이터 학습부는,

학습용 유전체 데이터를 입력하는 학습용 데이터 입력부와,

학습용 유전체 데이터 전처리하는 학습용 데이터 전처리부와,

기능 군(group)별로 학습용 데이터를 분리하는 데이터 분리부와,

교차 검증(5 fold cross-validation)을 위한 기능 군 별 학습(Classification)을 수행하는 기능군별 학습부를 포함하는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치.

청구항 4

제 1 항에 있어서, 외부 유전체 데이터 유효성 검사부는,

외부 유전체 데이터를 입력하는 외부 데이터 입력부와,

외부 유전체 데이터 전처리를 수행하는 외부 데이터 전처리부와,

전처리된 외부 유전체 데이터로 학습된 네트워크의 성능을 검증하여 네트워크 중 외부 유효성(external validation) 성능이 가장 높은 네트워크 선택을 하는 학습 네트워크 검증부를 포함하는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치.

청구항 5

삭제

청구항 6

제 1 항에 있어서, TCGA dataset 8개의 cancer type,

BRCA, DLBC, HNSC, KIRC, OV, PAAD, PRAD, SARC을 학습에 사용하여 외부데이터를 이용한 검증에서 정확도를 높

이는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치.

청구항 7

제 1 항에 있어서, 학습된 딥러닝 네트워크를 이용하여 암종 분류(cancer type classification)에 기여하는 각 유전자 및 기능별 유전자군의 기여도를 수치 표시를 포함하는 시각적으로 분석하기 위하여,

CAM(class activation mapping)을 이용한 분류 기여도 분석을 하는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치.

청구항 8

제 1 항에 있어서, 기능별 유전체 딥러닝 최적화를 통하여 학습에 필요한 gene 개수를 감소시키기 위하여,

C2H2 ZF, bZIP, bHLH, STAT, PAX5를 포함하는 TF의 하위 기능군 별로 암종 분류 기여도를 기준으로 이들의 조합 중 유전체 수 대비 가장 높은 정확도를 갖는 조합을 딥러닝을 이용한 암종 분류에 이용하는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치.

청구항 9

제 1 항에 있어서, 암종 분류의 정확도를 높이기 위하여,

기능(function)별 유전체 딥러닝으로 학습된 네트워크의 CAM(class activation mapping)의 강도값이 높은 순으로 gene을 선택한 후,

선택된 상위 10%의 gene들의 C2H2 ZF, bZIP, bHLH를 포함하는 기능별 group을 누적하여 상위에 있는 group들을 묶어서 다시 딥러닝 네트워크를 학습시켜서 가장 높은 정확도를 얻는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치.

청구항 10

유전체 데이터 학습부에서 학습용 유전체 데이터로 사용하기 위한 유전자 발현 데이터(gene expression data)를 입력받아 학습용 유전체 데이터를 전처리하고 교차 검증을 위한 기능 군 별 학습(Classification)을 수행하는 유전체 데이터 학습 단계;

외부 유전체 데이터 유효성 검사부에서 외부 유전체 데이터로 사용하기 위한 유전체 발현 데이터(gene expression data)를 입력받아 외부 유전체 데이터를 전처리하고, 전처리된 외부 유전체 데이터로 학습된 네트워크의 성능을 검증하여 외부 유효성(external validation) 성능이 가장 높은 네트워크를 선택하는 외부 유전체 데이터 유효성 검사 단계;를 포함하고,

유전체 발현 데이터(gene expression data)의 TF(transcription factor) 또는 세포막 유전체(cell membrane gene) 또는 microRNA의 특정 기능을 하는 군(group)을 개별적으로 딥러닝 학습하여 그 중 가장 분류 정확도 높은 군을 선택하여 딥러닝을 이용한 암종 분류(cancer type classification)의 대표 데이터로 사용하는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 방법.

청구항 11

제 10 항에 있어서, 유전체 데이터 학습 단계에서 입력되는 유전체 발현 데이터(gene expression data)는 TCGA(The Cancer Genome Atlas)이거나,

외부 유전체 데이터 유효성 검사 단계에서 입력되는 유전체 발현 데이터(gene expression data)는 ICGC(Internal Cancer Genome Consortium)인 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 방법.

청구항 12

삭제

청구항 13

제 10 항에 있어서, TCGA dataset 8개의 cancer type,

BRCA, DLBC, HNSC, KIRC, OV, PAAD, PRAD, SARC를 학습에 사용하여 외부데이터를 이용한 검증에서 정확도를 높이는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 방법.

청구항 14

제 10 항에 있어서, 학습된 딥러닝 네트워크를 이용하여 암종 분류(cancer type classification)에 기여하는 각 유전자 및 기능별 유전자군의 기여도를 수치 표시를 포함하는 시각적으로 분석하기 위하여,

CAM(class activation mapping)을 이용한 분류 기여도 분석을 하는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 방법.

청구항 15

제 10 항에 있어서, 기능별 유전체 딥러닝 최적화를 통하여 학습에 필요한 gene 개수를 감소시키기 위하여,

C2H2 ZF, bZIP, bHLH, STAT, PAX5를 포함하는 TF의 하위 기능군 별로 암종 분류 기여도를 기준으로 이들의 조합 중 유전체 수 대비 가장 높은 정확도를 갖는 조합을 딥러닝을 이용한 암종 분류에 이용하는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 방법.

청구항 16

제 10 항에 있어서, 암종 분류의 정확도를 높이기 위하여,

기능(function)별 유전체 딥러닝으로 학습된 네트워크의 CAM(class activation mapping)의 강도값이 높은 순으로 gene을 선택한 후,

선택된 상위 10%의 gene들의 C2H2 ZF, bZIP, bHLH를 포함하는 기능별 group을 누적하여 상위에 있는 group들을 묶어서 다시 딥러닝 네트워크를 학습시켜서 가장 높은 정확도를 얻는 것을 특징으로 하는 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 방법.

발명의 설명

기술 분야

[0001] 본 발명은 딥러닝을 이용한 암종 분류에 관한 것으로, 구체적으로 특정 기능을 하는 군(group)을 개별적으로 딥러닝 학습하여 그 중 가장 분류 정확도가 높은 몇 개의 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 한 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법에 관한 것이다.

배경 기술

[0002] 암종 분류는 암이 발견되었을 때, 그 원발 부위를 알아야 최적의 치료로 생존율을 높일 수 있다는 점에서 매우 중요하다.

[0003] 특히, 뼈에 발견된 암의 경우 그 원발 부위를 모르면 전이 여부를 알 수 없어, 적절한 치료시기와 치료법을 놓쳐 환자가 조기 사망할 위험이 있다.

[0004] 하지만, 종래 기술에서 암종 분류에 사용하는 딥러닝 방법들의 정확도가 높지 않았고, 딥러닝 시 사용되지 않은 외부 데이터에는 낮은 정확도를 보여주었다.

[0005] 유전체 데이터 전처리에 있어서는, 유전체 데이터를 이용한 종래 기술의 암종 분류(cancer type classification)의 전처리는 0에 가까운 값이나 변화량(variance)가 낮은 값을 제거하는 것 등이었는데, 작으나 유의미할지 모르는 데이터를 계량 통계값을 기준으로 삭제하는 것은 딥러닝 학습 정확도 하락의 원인이 될 수 있다.

[0006] 또한, 유전체 데이터 딥러닝에 있어서는, 유전체 순열을 이용하는 학습은 입력 데이터의 크기가 크고 네트워크의 파라미터 수가 너무 많아서 수만명 단위의 데이터가 있어야 정확도가 높다는 문제점이 있고, 유전체를 이미지화해서 학습하는 방식은 컨볼루션(convolution)을 통한 데이터 추상화로 인해 필요한 데이터 수가 작아진다는 장점이 있으나, 이미지 내의 유전체 배열이 실제 유전체의 배열을 반영하지 않기에 유효한 gene을 뽑아낼 수 없다는 문제점이 있다.

[0007] 따라서, 단순 계량 통계값을 이용해 전처리를 하여 발생하는 딥러닝 학습 정확도 하락 문제 및 유전체 데이터 딥러닝에 있어서 입력 데이터의 크기가 크고 네트워크의 파라미터 수가 너무 많이 필요한 문제를 해결할 수 있도록 하는 새로운 기술의 개발이 요구되고 있다.

선행기술문헌

특허문헌

- [0008] (특허문헌 0001) 대한민국 등록특허 제10-2044094호
- (특허문헌 0002) 대한민국 등록특허 제10-1950395호
- (특허문헌 0003) 대한민국 공개특허 제10-2018-0051333호

발명의 내용

해결하려는 과제

- [0009] 본 발명은 종래 기술의 딥러닝을 이용한 암종 분류 기술의 문제점을 해결하기 위한 것으로, 특정 기능을 하는 군(group)을 개별적으로 딥러닝 학습하여 그 중 가장 분류 정확도가 높은 몇 개의 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 한 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법을 제공하는데 그 목적이 있다.
- [0010] 본 발명은 가장 분류 정확도가 높은 몇 개의 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 하여 외부데이터를 이용한 검증에서 높은 검증 정확도를 갖도록 한 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법을 제공하는데 그 목적이 있다.
- [0011] 본 발명은 CAM(class activation mapping)을 이용한 분류 기여도 분석으로 학습된 딥러닝 네트워크를 이용하여 암종 분류에 기여하는 각 유전자 및 유전자군(기능별)의 기여도를 시각(수치)적으로 분석 가능하도록 한 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법을 제공하는데 그 목적이 있다.
- [0012] 본 발명은 TF의 하위 기능군(예: C2H2 ZF, bZIP, bHLH, STAT, PAX5) 별로 암종 분류의 기여도를 파악하여, 이들의 조합 중 유전체 수 대비 가장 높은 정확도를 갖는 조합을 딥러닝을 이용한 암종 분류에 사용하여 분류에 필요한 gene 수를 줄일 수 있도록 한 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법을 제공하는데 그 목적이 있다.
- [0013] 본 발명은 학습된 네트워크의 CAM의 강도 값을 이용하여 중요도 순(강도값이 높은 순)으로 gene을 선택한 후, 선택된 상위 gene들의 기능별 그룹(C2H2 ZF, bZIP, bHLH 등)을 누적하여 상위에 있는 그룹들을 묶어서 다시 딥러닝 네트워크를 학습시켜서 가장 높은 정확도를 얻을 수 있도록 하여 기능별 유전체 딥러닝 최적화가 가능하도록 한 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법을 제공하는데 그 목적이 있다.
- [0014] 본 발명은 높은 정확도를 갖는 gene data를 암종 분류의 대표 데이터로 사용할 수 있도록 하여 소수의 유전자 사용만으로도 암을 분류하는 진단 kit를 구성할 수 있도록 한 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법을 제공하는데 그 목적이 있다.
- [0015] 본 발명은 유전체의 이미지화 후 딥러닝으로 대표 데이터를 이미지화하여 분류하는 것에 의해 적은 환자수로도 높은 정확도를 얻을 수 있는 가볍고 빠른 네트워크를 획득할 수도 있도록 한 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법을 제공하는데 그 목적이 있다.
- [0016] 본 발명의 다른 목적들은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 또 다른 목적들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0017] 상기와 같은 목적을 달성하기 위한 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치는 학습용 유전체 데이터로 사용하기 위한 유전자 발현 데이터(gene expression data)를 입력받아 학습용 유전체 데이터를 전처리하고 교차 검증을 위한 기능 군 별 학습(Classification)을 수행하는 유전체 데이터 학습부;외부

유전체 데이터로 사용하기 위한 유전체 발현 데이터(gene expression data)를 입력받아 외부 유전체 데이터를 전처리하고, 전처리된 외부 유전체 데이터로 학습된 네트워크의 성능을 검증하여 외부 유효성(external validation) 성능이 가장 높은 네트워크를 선택하는 외부 유전체 데이터 유효성 검사부;를 포함하고, 특정 기능을 하는 유전체 데이터 군(group)을 개별적으로 딥러닝 학습하여 선택된 유전체 데이터 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 하는 것을 특징으로 한다.

[0018] 여기서, 유전체 데이터 학습부에 입력되는 유전체 발현 데이터(gene expression data)는 TCGA(The Cancer Genome Atlas)이거나, 외부 유전체 데이터 유효성 검사부에 입력되는 유전체 발현 데이터(gene expression data)는 ICGC(Internal Cancer Genome Consortium)인 것을 특징으로 한다.

[0019] 그리고 유전체 데이터 학습부는, 학습용 유전체 데이터를 입력하는 학습용 데이터 입력부와, 학습용 유전체 데이터 전처리하는 학습용 데이터 전처리부와, 기능 군(group)별로 학습용 데이터를 분리하는 데이터 분리부와, 교차 검증(5 fold cross-validation)을 위한 기능 군 별 학습(Classification)을 수행하는 기능군별 학습부를 포함하는 것을 특징으로 한다.

[0020] 그리고 외부 유전체 데이터 유효성 검사부는, 외부 유전체 데이터를 입력하는 외부 데이터 입력부와, 외부 유전체 데이터 전처리를 수행하는 외부 데이터 전처리부와, 전처리된 외부 유전체 데이터로 학습된 네트워크의 성능을 검증하여 네트워크 중 외부 유효성(external validation) 성능이 가장 높은 네트워크 선택을 하는 학습 네트워크 검증부를 포함하는 것을 특징으로 한다.

[0021] 그리고 유전체 발현 데이터(gene expression data)의 TF(transcription factor) 또는 세포막 유전체(cell membrane gene) 또는 microRNA의 특정 기능을 하는 군(group)을 개별적으로 딥러닝 학습하여 그 중 가장 분류 정확도 높은 군을 선택하여 딥러닝을 이용한 암종 분류(cancer type classification)의 대표 데이터로 사용하는 것을 특징으로 한다.

[0022] 그리고 TCGA dataset 8개의 cancer type, BRCA, DLBC, HNSC, KIRC, OV, PAAD, PRAD, SARC을 학습에 사용하여 외부데이터를 이용한 검증에서 정확도를 높이는 것을 특징으로 한다.

[0023] 학습된 딥러닝 네트워크를 이용하여 암종 분류(cancer type classification)에 기여하는 각 유전자 및 기능별 유전자군의 기여도를 수치 표시를 포함하는 시각적으로 분석하기 위하여, CAM(class activation mapping)을 이용한 분류 기여도 분석을 하는 것을 특징으로 한다.

[0024] 그리고 기능별 유전체 딥러닝 최적화를 통하여 학습에 필요한 gene 개수를 감소시키기 위하여, C2H2 ZF, bZIP, bHLH, STAT, PAX5를 포함하는 TF의 하위 기능군 별로 암종 분류 기여도를 기준으로 이들의 조합 중 유전체 수 대비 가장 높은 정확도를 갖는 조합을 딥러닝을 이용한 암종 분류에 이용하는 것을 특징으로 한다.

[0025] 그리고 암종 분류의 정확도를 높이기 위하여, 기능(function)별 유전체 딥러닝으로 학습된 네트워크의 CAM(class activation mapping)의 강도값이 높은 순으로 gene을 선택한 후, 선택된 상위 10%의 gene들의 C2H2 ZF, bZIP, bHLH를 포함하는 기능별 group을 누적하여 상위에 있는 group들을 묶어서 다시 딥러닝 네트워크를 학습시켜서 가장 높은 정확도를 얻는 것을 특징으로 한다.

[0026] 다른 목적을 달성하기 위한 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 방법은 학습용 유전체 데이터로 사용하기 위한 유전자 발현 데이터(gene expression data)를 입력받아 학습용 유전체 데이터를 전처리하고 교차 검증을 위한 기능 군 별 학습(Classification)을 수행하는 유전체 데이터 학습 단계;외부 유전체 데이터로 사용하기 위한 유전체 발현 데이터(gene expression data)를 입력받아 외부 유전체 데이터를 전처리하고, 전처리된 외부 유전체 데이터로 학습된 네트워크의 성능을 검증하여 외부 유효성(external validation) 성능이 가장 높은 네트워크를 선택하는 외부 유전체 데이터 유효성 검사 단계;를 포함하고, 특정 기능을 하는 유전체 데이터 군(group)을 개별적으로 딥러닝 학습하여 선택된 유전체 데이터 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 하는 것을 특징으로 한다.

[0027] 여기서, 유전체 데이터 학습 단계에서 입력되는 유전체 발현 데이터(gene expression data)는 TCGA(The Cancer Genome Atlas)이거나, 외부 유전체 데이터 유효성 검사 단계에서 입력되는 유전체 발현 데이터(gene expression data)는 ICGC(Internal Cancer Genome Consortium)인 것을 특징으로 한다.

[0028] 그리고 유전체 발현 데이터(gene expression data)의 TF(transcription factor) 또는 세포막 유전체(cell membrane gene) 또는 microRNA의 특정 기능을 하는 군(group)을 개별적으로 딥러닝 학습하여 그 중 가장 분류 정확도 높은 군을 선택하여 딥러닝을 이용한 암종 분류(cancer type classification)의 대표 데이터로 사용하는

것을 특징으로 한다.

- [0029] 그리고 TCGA dataset 8개의 cancer type, BRCA, DLBC, HNSC, KIRC, OV, PAAD, PRAD, SARC를 학습에 사용하여 외부데이터를 이용한 검증에서 정확도를 높이는 것을 특징으로 한다.
- [0030] 그리고 학습된 딥러닝 네트워크를 이용하여 암종 분류(cancer type classification)에 기여하는 각 유전자 및 기능별 유전자군의 기여도를 수치 표시를 포함하는 시각적으로 분석하기 위하여, CAM(class activation mapping)을 이용한 분류 기여도 분석을 하는 것을 특징으로 한다.
- [0031] 그리고 기능별 유전체 딥러닝 최적화를 통하여 학습에 필요한 gene 개수를 감소시키기 위하여, C2H2 ZF, bZIP, bHLH, STAT, PAX5를 포함하는 TF의 하위 기능군 별로 암종 분류 기여도를 기준으로 이들의 조합 중 유전체 수 대비 가장 높은 정확도를 갖는 조합을 딥러닝을 이용한 암종 분류에 이용하는 것을 특징으로 한다.
- [0032] 그리고 암종 분류의 정확도를 높이기 위하여, 기능(function)별 유전체 딥러닝으로 학습된 네트워크의 CAM(class activation mapping)의 강도값이 높은 순으로 gene을 선택한 후, 선택된 상위 10%의 gene들의 C2H2 ZF, bZIP, bHLH를 포함하는 기능별 group을 누적하여 상위에 있는 group들을 묶어서 다시 딥러닝 네트워크를 학습시켜서 가장 높은 정확도를 얻는 것을 특징으로 한다.

발명의 효과

- [0033] 이상에서 설명한 바와 같은 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법은 다음과 같은 효과가 있다.
- [0034] 첫째, 특정 기능을 하는 군(group)을 개별적으로 딥러닝 학습하여 그 중 가장 분류 정확도가 높은 몇 개의 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 한다.
- [0035] 둘째, 가장 분류 정확도가 높은 몇 개의 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 하여 외부데이터를 이용한 검증에서 높은 검증 정확도를 갖도록 한다.
- [0036] 셋째, CAM(class activation mapping)을 이용한 분류 기여도 분석으로 학습된 딥러닝 네트워크를 이용하여 암종 분류에 기여하는 각 유전자 및 유전자군(기능별)의 기여도를 시각(수치)적으로 분석 가능하도록 한다.
- [0037] 넷째, TF의 하위 기능군(예: C2H2 ZF, bZIP, bHLH, STAT, PAX5) 별로 암종 분류의 기여도를 파악하여, 이들의 조합 중 유전체 수 대비 가장 높은 정확도를 갖는 조합을 딥러닝을 이용한 암종 분류에 사용하여 분류에 필요한 gene 수를 줄일 수 있도록 한다.
- [0038] 다섯째, 학습된 네트워크의 CAM의 강도 값을 이용하여 중요도 순(강도값이 높은 순)으로 gene을 선택한 후, 선택된 상위 gene들의 기능별 그룹(C2H2 ZF, bZIP, bHLH 등)을 누적하여 상위에 있는 그룹들을 묶어서 다시 딥러닝 네트워크를 학습시켜서 가장 높은 정확도를 얻을 수 있도록 하여 기능별 유전체 딥러닝 최적화가 가능하도록 한다.
- [0039] 여섯째, 높은 정확도를 갖는 gene data를 암종 분류의 대표 데이터로 사용할 수 있도록 하여 소수의 유전자 사용만으로도 암을 분류하는 진단 kit를 구성할 수 있도록 한다.
- [0040] 일곱째, 유전체의 이미지화 후 딥러닝으로 대표 데이터를 이미지화하여 분류하는 것에 의해 적은 환자수로도 높은 정확도를 얻을 수 있는 가볍고 빠른 네트워크를 획득할 수도 있도록 한다.

도면의 간단한 설명

- [0041] 도 1은 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치의 구성도
- 도 2는 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 방법을 나타낸 플로우 차트
- 도 3은 TF(Transcription factor)로만 학습시킨 TCGA 암종 분류 결과 테이블
- 도 4는 학습된 네트워크의 평균 CAM(class activation map)
- 도 5는 TF에서 각 암종별로 CAM에서 높은 강도값을 보이는 gene과 그 기능별 그룹 테이블

발명을 실시하기 위한 구체적인 내용

- [0042] 이하, 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법의 바람직한 실시 예에 관

하여 상세히 설명하면 다음과 같다.

- [0043] 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법의 특징 및 이점들은 이하에서의 각 실시 예에 대한 상세한 설명을 통해 명백해질 것이다.
- [0044] 도 1은 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치의 구성도이다.
- [0045] 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법은 특정 기능을 하는 군(group)을 개별적으로 딥러닝 학습하여 그 중 가장 분류 정확도가 높은 몇 개의 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 한 것이다.
- [0046] 이를 위하여, 본 발명은 가장 분류 정확도가 높은 몇 개의 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 하여 외부데이터를 이용한 검증에서 높은 검증 정확도를 갖도록 하는 구성을 포함할 수 있다.
- [0047] 본 발명은 CAM(class activation mapping)을 이용한 분류 기여도 분석으로 학습된 딥러닝 네트워크를 이용하여 암종 분류에 기여하는 각 유전자 및 유전자군(기능별)의 기여도를 시각(수치)적으로 분석 가능하도록 하는 구성을 포함할 수 있다.
- [0048] 본 발명은 TF의 하위 기능군(예: C2H2 ZF, bZIP, bHLH, STAT, PAX5) 별로 암종 분류의 기여도를 파악하여, 이들의 조합 중 유전체 수 대비 가장 높은 정확도를 갖는 조합을 딥러닝을 이용한 암종 분류에 사용하여 분류에 필요한 gene 수를 줄일 수 있도록 하는 구성을 포함할 수 있다.
- [0049] 본 발명은 학습된 네트워크의 CAM의 강도 값을 이용하여 중요도 순(강도값이 높은 순)으로 gene을 선택한 후, 선택된 상위 gene들의 기능별 그룹(C2H2 ZF, bZIP, bHLH 등)을 누적하여 상위에 있는 그룹들을 묶어서 다시 딥러닝 네트워크를 학습시켜서 가장 높은 정확도를 얻을 수 있도록 하여 기능별 유전체 딥러닝 최적화가 가능하도록 하는 구성을 포함할 수 있다.
- [0050] 본 발명은 높은 정확도를 갖는 gene data를 암종 분류의 대표 데이터로 사용할 수 있도록 하여 소수의 유전자 사용만으로도 암을 분류하는 진단 kit를 구성할 수 있도록 하는 구성을 포함할 수 있다.
- [0051] 본 발명은 유전체의 이미지화 후 딥러닝으로 대표 데이터를 이미지화하여 분류하는 것에 의해 적은 환자수로도 높은 정확도를 얻을 수 있는 가볍고 빠른 네트워크를 획득할 수도 있도록 하는 구성을 포함할 수 있다.
- [0052] 이하의 설명에서 '유전체 데이터'는 TCGA(The Cancer Genome Atlas) 및 ICGC(Internal Cancer Genome Consortium)와 같이 대규모 암 유전체 연구 프로젝트를 통해 각기 다른 여러 기술적 플랫폼을 이용하여 생산된 다양한 형태의 데이터일 수 있다.
- [0053] 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치는 도 1에서와 같이, 학습용 유전체 데이터(TCGA 등의 유전자 발현 데이터(gene expression data))를 입력받아 학습용 유전체 데이터를 전처리(\log_2X+1)하고 교차 검증(5 fold cross-validation)을 위한 기능 군 별 학습(Classification)을 수행하는 유전체 데이터 학습부(100)와, 외부 유전체 데이터(ICGC 등의 유전체 발현 데이터(gene expression data))를 입력받아 외부 유전체 데이터를 전처리(\log_2X+1)하고 전처리된 외부 유전체 데이터로 학습된 네트워크의 성능을 검증하여 가장 external validation 성능이 높은 네트워크를 선택하는 외부 유전체 데이터 유효성 검사부(200)를 포함한다.
- [0054] 여기서, 유전체 데이터 학습부(100)는 학습용 유전체 데이터(TCGA 등의 유전자 발현 데이터(gene expression data))를 입력하는 학습용 데이터 입력부(10)와, 학습용 유전체 데이터 전처리(\log_2X+1)하는 학습용 데이터 전처리부(20)와, transcription factor, cell membrane protein 등의 기능 군(group)별로 학습용 데이터를 분리하는 데이터 분리부(30)와, 교차 검증(5 fold cross-validation)을 위한 기능 군 별 학습(Classification)을 수행하는 기능군별 학습부(40)를 포함한다.
- [0055] 그리고 외부 유전체 데이터 유효성 검사부(200)는 외부 유전체 데이터(ICGC 등의 유전체 발현 데이터(gene expression data))를 입력하는 외부 데이터 입력부(50)와, 외부 유전체 데이터 전처리(\log_2X+1)를 수행하는 외부 데이터 전처리부(60)와, 전처리된 외부 유전체 데이터로 학습된 네트워크의 성능을 검증하여 5개의 네트워크 중 가장 external validation 성능이 높은 네트워크를 선택하는 학습 네트워크 검증부(70)를 포함한다.
- [0056] 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 방법을 구체적으로 설명하면 다음과 같다.
- [0057] 도 2는 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 방법을 나타낸 플로우 차트이다.

- [0058] 먼저, 학습용 유전체 데이터(TCGA 등의 유전자 발현 데이터(gene expression data))를 입력받는다.(S201)
- [0059] 이어, 입력된 학습용 유전체 데이터의 전처리(\log_2X+1)를 수행한다.(S202)
- [0060] 그리고 transcription factor, cell membrane protein 등의 기능 군(group)별로 학습용 데이터를 분리한다.(S203)
- [0061] 이어, 교차 검증(5 fold cross-validation)을 위한 기능군별 학습(Classification)을 수행한다.(S204)
- [0062] 그리고 외부 유전체 데이터 유효성 검사를 위하여, 외부 유전체 데이터(ICGC 등의 유전체 발현 데이터(gene expression data))를 입력받는다.(S205)
- [0063] 이어, 외부 유전체 데이터 전처리(\log_2X+1)를 수행하고(S206), 전처리된 외부 유전체 데이터로 학습된 네트워크의 성능을 검증하여 네트워크 중 가장 external validation 성능이 높은 네트워크 선택을 한다.(S206)
- [0064] 본 발명에 따른 기능(function)별 유전체 딥러닝에 관하여 구체적으로 설명하면 다음과 같다.
- [0065] 기능별 학습은 단순 계량 통계값을 이용해 전처리를 하는 문제에서 벗어나, 유전체 발현 데이터(gene expression data)의 TF(transcription factor), 세포막 유전체(cell membrane gene), microRNA 등의 특정 기능을 하는 군(group)을 개별적으로 딥러닝 학습하여 그 중 가장 분류 정확도가 높은 몇 개의 군을 딥러닝을 이용한 암종 분류(cancer type classification)의 대표 데이터로 사용할 수 있다.
- [0066] 그리고 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법은 높은 검증 정확도를 갖는다.
- [0067] TF만(1468개, TCGA dataset 8개 cancer type: BRCA, DLBC, HNSC, KIRC, OV, PAAD, PRAD, SARC)을 학습에 사용하였을 때 평균 0.99의 정확도를 보였고, 외부데이터를 이용한 검증인 external validation(ICGC dataset)에서도 약 0.96의 높은 정확도를 보인다.
- [0068] 도 3은 TF(Transcription factor)로만 학습시킨 TCGA 암종 분류 결과 테이블 및 외부데이터를 이용한 검증인 external validation(ICGC dataset) 특성 테이블이다.
- [0069] 특히, 본 발명은 CAM(class activation mapping)을 이용한 분류 기여도 분석으로 학습된 딥러닝 네트워크를 이용하여 암종 분류(cancer type classification)에 기여하는 각 유전자 및 유전자군(기능별)의 기여도를 시각(수치)적으로 분석 가능하도록 한다.
- [0070] 도 4는 학습된 네트워크의 평균 CAM(class activation map)이다.
- [0071] 또한, 본 발명은 기능별 유전체 딥러닝 최적화를 통하여 학습에 필요한 gene 개수를 감소시킬 수 있다.
- [0072] 도 5 및 표 1에서와 같이, TF의 하위 기능군(예: C2H2 ZF, bZIP, bHLH, STAT, PAX5) 별로 cancer classification의 기여를 파악하여, 이들의 조합 중 유전체 수 대비 가장 높은 accuracy를 보이는 조합을 딥러닝을 이용한 cancer type classification에 이용하여 분류에 필요한 gene 수를 줄일 수 있다.
- [0073] 도 5는 TF에서 각 암종별로 CAM에서 높은 강도값을 보이는 gene과 그 기능별 그룹 테이블이다.

표 1

function_id	count
C2H2 ZF	287
Unknown	140
bZIP	133
bHLH	106
HMG/Sox	39
STAT	38
Homeodomain	35
Forkhead	34
Nuclear receptor	34
Ets	28
AT hook	26
Rel	24
SMAD	24
CxxC	20
MBD	17
IRF	16
MADS box	14
GATA	12
Myb/SANT	12
ARID/BRIGHT	11
E2F	11
MBD; AT hook	11

[0074]

[0075]

본 발명은 TF에서 학습에 필요한 gene 개수 감소시킨다.

[0076]

기능(function)별 유전체 딥러닝으로 학습된 네트워크의 class activation mapping(CAM)의 강도 값을 이용하여 중요도 순(강도값이 높은 순)으로 gene을 선택한 후, 선택된 상위 10%의 gene들의 기능별 group(C2H2 ZF, bZIP, bHLH 등)을 누적하여 상위에 있는 group들을 묶어서 다시 deep learning 네트워크를 학습시켜서 가장 높은 정확도를 얻을 수 있다.

[0077]

표 2에서와 같이, TF 중 C2H2 ZF (272개의 유전자)만으로 0.97의 높은 accuracy를 갖는 gene data를 암종 분류의 대표 데이터로 사용할 수 있다.

[0078]

따라서 소수의 유전자만으로 암을 분류하는 진단 kit로 사용될 수 있도록 한다.

[0079]

C2H2 ZF, bZIP, bHLH 3개 기능군의 조합(약 800개의 유전자)으로 0.98의 높은 accuracy를 갖는 gene data를 암종 분류의 대표 데이터로 사용할 수 있으며, 따라서 암종 발견 시 이 800개의 gene data를 사용하여 신속하고 정확하게 암 검진이 가능하도록 한다.

표 2

		Accuracy		Accuracy
net_TF660index_ver1 (C2H2 ZF)	validation	0.9666	test	0.9674
net_TF809index_ver (C2H2 ZF + bZIP + bHLH)	validation	0.9826	test	0.9819

[0080]

[0081]

따라서 CH2H ZF gene 단독이나, 다른 기능군인 bZIP, bHLH, STAT, PAX5 등과 함께 조합되어 암 분류를 위한 신속 진단 kit로 사용가능하다.

[0082]

또한 암종에 따라 TF 기능별 그룹을 2개, 3개, 4개...씩 무작위로 조합하여 가장 높은 accuracy를 보이는 그룹을 선택하는 것도 가능하다.

[0083] 그리고 유전체의 이미지화 후 딥러닝으로, 대표 데이터를 이미지화하여 classification함으로써 적은 환자수로도 높은 정확도를 얻을 수 있는 가볍고 빠른 네트워크를 획득할 수도 있다.

[0084] 이상에서 설명한 본 발명에 따른 유전체 기능별 딥러닝을 이용한 암종 분류를 위한 장치 및 방법은 특정 기능을 하는 군(group)을 개별적으로 딥러닝 학습하여 그 중 가장 분류 정확도가 높은 몇 개의 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 한 것이다.

[0085] 본 발명은 가장 분류 정확도가 높은 몇 개의 군을 딥러닝을 이용한 암종 분류의 대표 데이터로 사용할 수 있도록 하여 외부데이터를 이용한 검증에서 높은 검증 정확도를 갖도록 한 것이다.

[0086] 이상에서의 설명에서와 같이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 본 발명이 구현되어 있음을 이해할 수 있을 것이다.

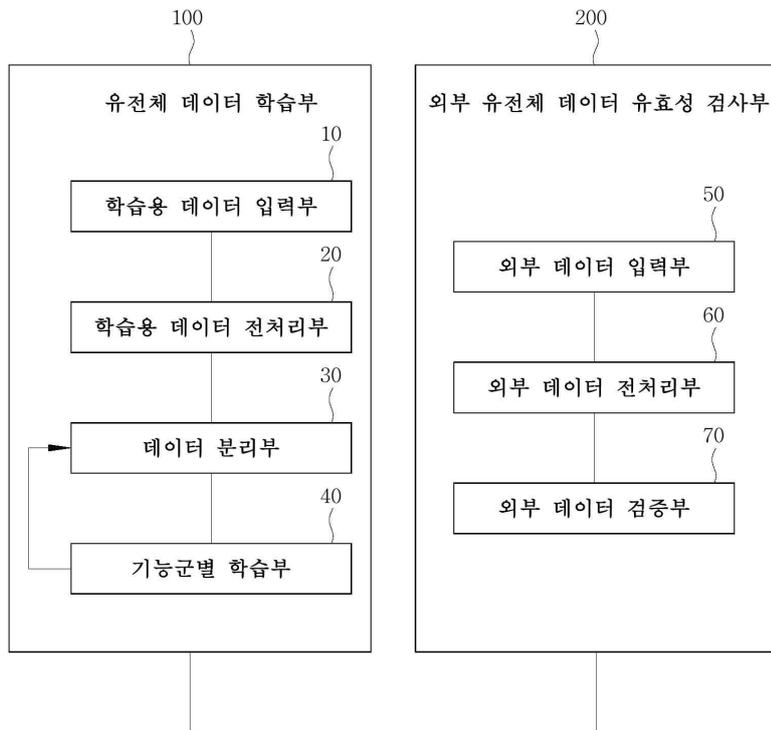
[0087] 그러므로 명시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 하고, 본 발명의 범위는 전술한 설명이 아니라 특허청구 범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

부호의 설명

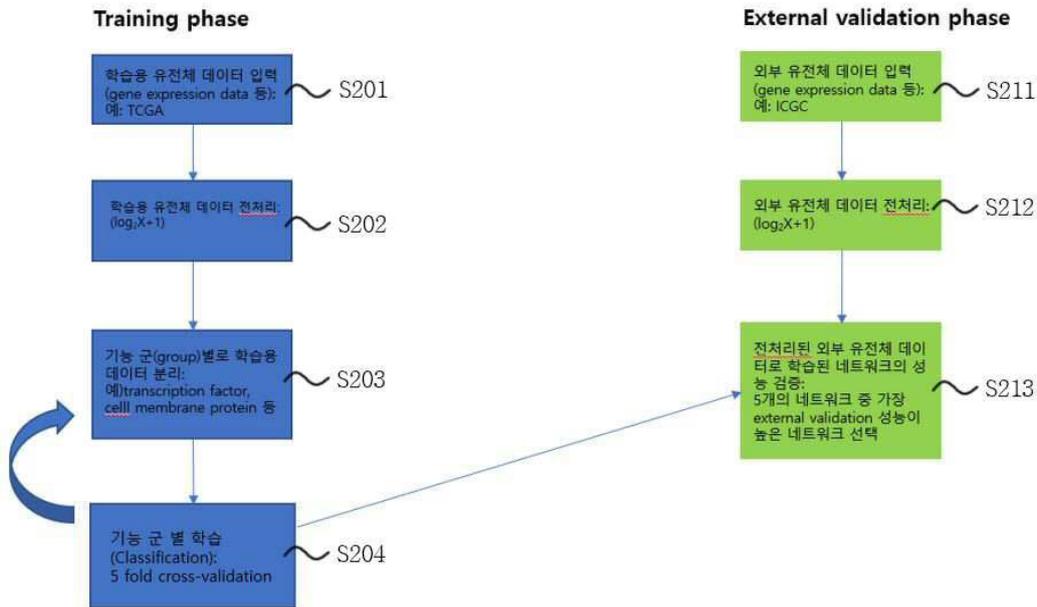
- [0088] 100. 유전체 데이터 학습부
- 200. 외부 유전체 데이터 유효성 검사부

도면

도면1



도면2



도면3

TCGA

Training set(0.8) : Validation set(0.2)

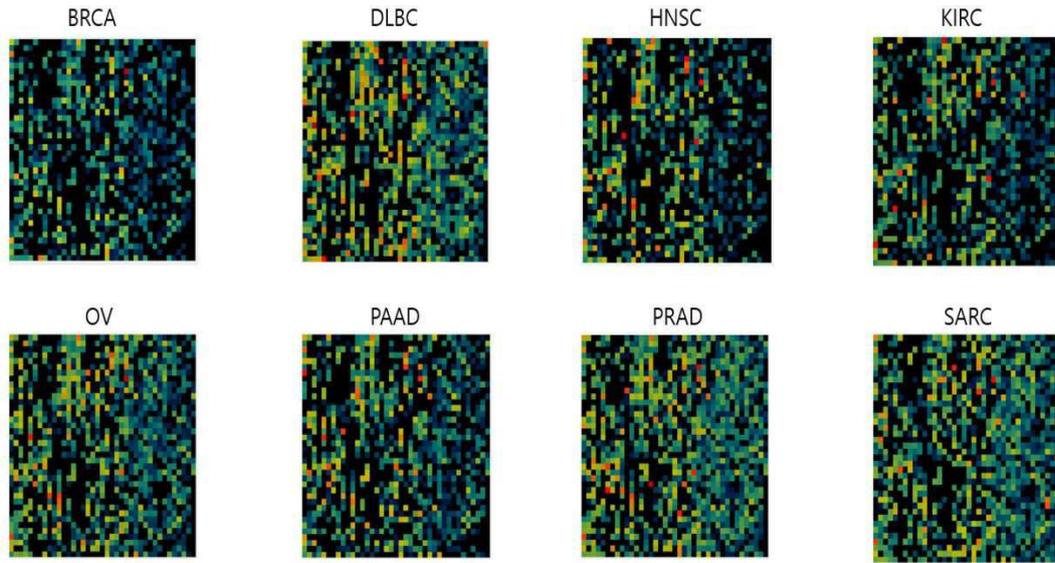
	<u>TCGA Accuracy Validation</u>	<u>Precision</u>	<u>Recall</u>	<u>F score</u>
1net	0.9913	0.99	0.9761	0.983
2net	0.9927	0.9947	0.9875	0.9911
3net	0.9942	0.995	0.9899	0.9925
4net	0.9927	0.9953	0.9796	0.9874
5net	0.9868	0.9869	0.9564	0.9714

ICGC

Test set(All)

	<u>ICGC Accuracy Test</u>	<u>Precision</u>	<u>Recall</u>	<u>F score</u>
1net	0.9583	0.9686	0.9621	0.9653
2net	0.8895	0.9198	0.9159	0.9179
3net	0.9891	0.9871	0.9874	0.9873
4net	0.9855	0.9836	0.9839	0.9837
5net	0.9837	0.9822	0.9812	0.9817

도면4



도면5

BRCA	function	DLBC	function	HNSC	function	KIRC	function	OV	function	PAAD	function	PRAD	function	SARC	function
XPB1	bZIP	CHCHD3	Unknown	STAT1	STAT	NFE2L1	bZIP	YBX1	CSD	AHR	bHLH	NKX3-1	Homeodomain	YBX1	CSD
AEBP1	Unknown	ETS1	Ets	TP63	p53	EPAS1	bHLH	ELF3	Ets; AT hook	ELF3	Ets; AT hook	XPB1	bZIP	PAX7	Homeodomain; Paired box
GATA3	GATA	IRF8	IRF	HIF1A	bHLH	TSC22D1	Unknown	ATF4	bZIP	YBX1	CSD	FOXA1	Forkhead	STAT6	STAT
STAT1	STAT	STAT6	STAT	YBX1	CSD	YBX1	CSD	HMGAI	AT hook	HIF1A	bHLH	TSC22D1	Unknown	ATF4	bZIP
BHLHE40	bHLH	STAT1	STAT	STAT3	STAT	ATF4	bZIP	JUN	bZIP	STAT3	STAT	SPODF	Ets	FOX	bZIP
TFPI1	GATA	DMMT1	Cx/C	AHR	bHLH	FOS	bZIP	STAT1	STAT	FOS	bZIP	ATF4	bZIP	ZNF704	C2H2 ZF
ATF4	bZIP	PAX5	Paired box	JUNB	bZIP	KLf6	C2H2 ZF	TSC22D1	Unknown	KLf6	C2H2 ZF	HDXB13	Homeodomain	BCL118	C2H2 ZF
SREBF1	bHLH	POU2AF1	Unknown	IRF6	IRF	JUN	bZIP	FOSL2	bZIP	STAT1	STAT	FOS	bZIP	ZNF703	C2H2 ZF
YBX1	CSD	TCF3	bHLH	FOS	bZIP	ZNF395	C2H2 ZF	PAX8	Paired box	BHLHE40	bHLH	JUN	bZIP	TCF3	bHLH
NFE2L1	bZIP	NR1H2	Nuclear receptor	NFE2L1	bZIP	STAT6	STAT	JUNB	bZIP	SKIL	Unknown	YBX1	CSD	ZEB2	C2H2 ZF; Homeodomain
MAZ	C2H2 ZF	ZNF90	C2H2 ZF	ETS2	Ets	AEBP1	Unknown	NFE2L1	bZIP	AEBP1	Unknown	CREB3L4	bZIP	NFIX	SMAD
FOXO1	Forkhead	AKNA	AT hook	EPAS1	bHLH	MAF	bZIP	PA2G4	Unknown	KLf5	C2H2 ZF	HMG208	HMG/Sox	MYC	bHLH
PRK1	Homeodomain	KZF1	C2H2 ZF	JUN	bZIP	BHLHE40	bHLH	MA2	C2H2 ZF	NFE2L1	bZIP	TRK3	T-box	CIC	HMG/Sox
FOS	bZIP	HIF1A	bHLH	PA2G4	Unknown	EGR1	C2H2 ZF	JUND	bZIP	FOSL2	bZIP	NFIX	SMAD	ADNP	Homeodomain
JUN	bZIP	POU2F2	Homeodomain; POU/STAT2	STAT	ETS1	Ets	STAT3	STAT	JUNB	bZIP	ZKSCAN1	C2H2 ZF	KDMA3	Cx/C	
STAT3	STAT	NCOA3	bHLH	KLf5	C2H2 ZF	STAT1	STAT	STAT3	STAT	ZKSCAN1	C2H2 ZF	NFE2L1	bZIP	STAT2	STAT
ESR1	Nuclear receptor	HMGAI	AT hook	ATF4	bZIP	HMGN3	HMG/Sox	XPB1	bZIP	TSC22D1	Unknown	NR4A1	Nuclear receptor	STAT3	STAT
RUNX1	Runt	SREBF2	bHLH	XPB1	bZIP	PAX8	Paired box	NR2F6	Nuclear receptor	SP1	C2H2 ZF	CREB3L1	bZIP	SABF	Unknown
KDMA2	Cx/C	USF2	bHLH	BHLHE40	bHLH	STAT3	STAT	STAT6	STAT	ZBTB38	C2H2 ZF	EGR1	C2H2 ZF	BAZZA	MBD; AT hook
STAT6	STAT	TCF4	bHLH	BHF	Ets	CUX1	CUT; Homeodomain	CXCC5	Cx/C	XPB1	bZIP	JUNB	bZIP	KLf6	C2H2 ZF
ESR1	C2H2 ZF	STAT3	STAT	FOSL2	bZIP	KLf9	C2H2 ZF	ZNF146	C2H2 ZF	EPAS1	bHLH	STAT3	STAT	CGGBP1	Unknown
TSC22D1	Unknown	YBX1	CSD	HMGAI	AT hook	FOSL2	bZIP	KDMA2	Cx/C	NFAT5	Rel	STAT6	STAT	ZNF83	C2H2 ZF
SRAP	AT hook	MBD2	MBD	KDMA5	ARID/BRIGHT	PA2G4	Unknown	BHLHE41	bHLH	ATF4	bZIP	ELK4	Ets	AEBP1	Unknown
JUNB	bZIP	GF3A	C2H2 ZF	MXD1	bHLH	SP1	C2H2 ZF	KLf6	C2H2 ZF	BTF	Unknown	SREBF2	bHLH	SKI	Unknown
ADNP	Homeodomain	SP100	SAND	BACH1	bZIP	ETS2	Ets	MEIS1	Homeodomain	KDMA5	ARID/BRIGHT	GATA2	GATA	UBP1	Grainyhead
SOX4	HMG/Sox	ESR2	Nuclear receptor	KDMA2	Cx/C	GPBP1	Unknown	USF2	bHLH	FOXO3	Forkhead	SREBF1	bHLH	PBX1	Homeodomain
ZNF24	C2H2 ZF	RFX5	RFX	KLf13	C2H2 ZF	NFE2L2	bZIP	YY1	C2H2 ZF	BHF	Ets	AEBP1	Unknown	ERF	Ets
REPN1	C2H2 ZF	BCL6	C2H2 ZF	NCOA3	bHLH	CREB2	bZIP	HIF1A	bHLH	SP3	C2H2 ZF	AR	Nuclear receptor	JUN	bZIP
KLf6	C2H2 ZF	MA2	C2H2 ZF	AEBP1	Unknown	NFIB	SMAD	GPBP1L1	Unknown	JUN	bZIP	GRHL2	Grainyhead	TSC22D1	Unknown
KDMA5	ARID/BRIGHT	DMT1	Myb/SANT	SP1	C2H2 ZF	MXI1	bHLH	RBCK1	Unknown	MBNL2	CCCH ZF	REPN1	C2H2 ZF	ZKSCAN1	C2H2 ZF
SP1	C2H2 ZF	MEF2C	MADS box	KLf6	C2H2 ZF	GF3A	C2H2 ZF	ADNP	Homeodomain	STAT6	STAT	JUND	STAT	GPBP1	Unknown
SPODF	Ets	BAZZA	MBD; AT hook	ARNTL2	bHLH	CXCC5	Cx/C	SOX17	HMG/Sox	HMGAI	AT hook	BHF	Ets	ZNF532	C2H2 ZF
SREBF2	bHLH	SABF	Unknown	SP3	C2H2 ZF	XPB1	bZIP	ATF68	bZIP	BBX	HMG/Sox	ADNP	Homeodomain	JUND	bZIP
BAZZA	MBD; AT hook	ATF4	bZIP	BNCl	C2H2 ZF	USF2	bHLH	HSP1	HSF	KLf3	C2H2 ZF	STAT1	STAT	PBX2	Homeodomain
JUND	bZIP	BTF	Unknown	CUX1	CUT; Homeodomain	HIF1A	bHLH	PBX1	Homeodomain	ETS2	Ets	ZNF24	C2H2 ZF	NFE2L1	bZIP
ELF3	Ets; AT hook	MECOM	C2H2 ZF	NFE2L2	bZIP	FOXJ3	Forkhead	PBX2	Homeodomain	CUX1	CUT; Homeodomain	ZBTB78	C2H2 ZF	AKAP8L	C2H2 ZF
STAT2	STAT	SP140	SAND	MYC	bHLH	POU3F3	Homeodomain; POU/SCAP	AT hook	TEAD1	TEA	HIF1A	bHLH	SP1	C2H2 ZF	
PA2G4	Unknown	MYBL2	Myb/SANT	SREBF2	bHLH	YY1	C2H2 ZF	REPN1	C2H2 ZF	SPEN	Unknown	SGSM2	BED ZF	SRAP	AT hook
HMG208	HMG/Sox	KDMA2	Cx/C	STAT6	STAT	GPBP1L1	Unknown	WT1	C2H2 ZF	YY1	C2H2 ZF	ZBTB10	C2H2 ZF	DMT1	Myb/SANT
GPBP1L1	Unknown	MBD4	MBD	SKIL	Unknown	NR3C1	Nuclear receptor	ZNF503	C2H2 ZF	ADNP	Homeodomain	SRAP	AT hook	SREBF2	bHLH
GLYR1	AT hook	CGGBP1	Unknown	SP100	SAND	RBCK1	Unknown	SP1	C2H2 ZF	ZNF217	C2H2 ZF	CUX1	CUT; Homeodomain	TCF12	bHLH
ATF68	bZIP	NFE2L1	bZIP	MAF	bZIP	MBNL2	CCCH ZF	HMGN3	HMG/Sox	STAT2	STAT	NFIX	SMAD	PA2G4	Unknown