



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0078884
(43) 공개일자 2024년06월04일

(51) 국제특허분류(Int. Cl.)
G16H 50/20 (2018.01) G16H 30/40 (2018.01)
G16H 50/50 (2018.01)
(52) CPC특허분류
G16H 50/20 (2018.01)
G16H 30/40 (2018.01)
(21) 출원번호 10-2022-0161531
(22) 출원일자 2022년11월28일
심사청구일자 2022년11월28일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
정승원
서울특별시 서초구 서초대로65길 13-10, 113동 1603호(서초동, 서초래미안아파트)
쉬썬신
서울특별시 동대문구 안암로18가길 27, 하늘정원 601호(제기동)
(뒷면에 계속)
(74) 대리인
송인호, 윤형근, 최관락

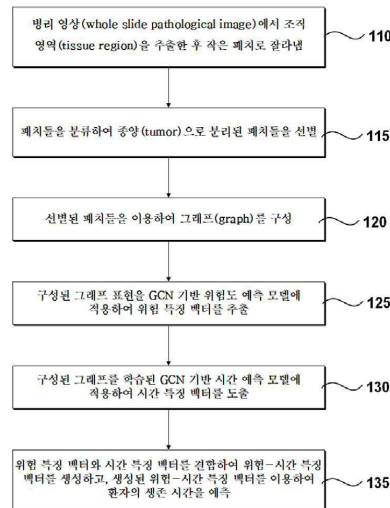
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 병리 영상을 이용한 생존 시간 예측 방법 및 그 장치

(57) 요약

병리 영상을 이용한 생존 시간 예측 방법 및 그 장치가 개시된다. 병리 영상을 이용한 생존 시간 예측 방법은, (a) 병리 영상을 전처리하여 조직(tissue) 영역을 각각 추출한 후 각각 패치들로 잘라내는 단계; (b) 상기 각각의 패치들을 그래프 구조로 구성하는 단계; (c) 상기 구성된 그래프 구조를 학습된 GCN 기반 위험도 예측 모델에 적용하여 위험 특징 벡터를 도출하는 단계; (d) 상기 구성된 그래프 구조를 학습된 GCN 기반 시간 예측 모델에 적용하여 시간 특징 벡터를 도출하는 단계; 및 (e) 상기 위험 특징 벡터와 상기 시간 특징 벡터를 결합한 위험-시간 특징 벡터를 이용하여 예측된 생존 시간을 출력하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

G16H 50/50 (2018.01)

(72) 발명자

임서훈

경기도 김포시 고촌읍 수기로 67-53, 219동 1402
호(수기마을힐스테이트2단지)

루위청

서울특별시 동대문구 무학로49길 84, 연희빌 301
호(용두동)

이 발명을 지원한 국가연구개발사업

과제고유번호 1465036483

과제번호 HI21C0940130022

부처명 보건복지부

과제관리(전문)기관명 한국보건산업진흥원

연구사업명 디지털 병리 기반의 암 전문 AI 분석 솔루션 개발

연구과제명 디지털 병리 및 인공지능 기반 다중 장기 지능형 형태 분석 및 분자변이예측 소프트

웨어 개발

기 여 율 1/1

과제수행기관명 고려대학교산학협력단

연구기간 2022.01.01 ~ 2022.12.31

명세서

청구범위

청구항 1

- (a) 병리 영상을 전처리하여 조직(tissue) 영역을 각각 추출한 후 각각 패치들로 잘라내는 단계;
- (b) 상기 각각의 패치들을 그래프 구조로 구성하는 단계;
- (c) 상기 구성된 그래프 구조를 학습된 GCN 기반 위험도 예측 모델에 적용하여 위험 특징 벡터를 도출하는 단계;
- (d) 상기 구성된 그래프 구조를 학습된 GCN 기반 시간 예측 모델에 적용하여 시간 특징 벡터를 도출하는 단계; 및
- (e) 상기 위험 특징 벡터와 상기 시간 특징 벡터를 결합한 위험-시간 특징 벡터를 이용하여 예측된 생존 시간을 출력하는 단계를 포함하는 병리 영상을 이용한 생존 시간 예측 방법.

청구항 2

제1 항에 있어서,

상기 (b) 단계는,

상기 각각의 패치를 기학습된 중앙 분류기에 적용하여 중앙에 해당하는 패치들을 선별하는 단계;

상기 선별된 각 패치들의 특징값을 각각 추출하고, 각 패치들의 특징값을 그래프의 꼭지점으로 각각 설정하는 단계; 및

상기 각각의 패치 중심 위치 사이의 유클리드 거리를 기반으로 K-최근접 이웃 알고리즘을 적용하여 꼭지점 간의 에지를 각각 설정하여 상기 그래프 구조로 구성하는 단계를 포함하는 것을 특징으로 하는 병리 영상을 이용한 생존 시간 예측 방법.

청구항 3

제1 항에 있어서,

상기 (e) 단계는,

상기 위험 특징 벡터의 순서 정보를 고려하여 상기 시간 특징 벡터가 보정되는 것을 특징으로 하는 병리 영상을 이용한 생존 시간 예측 방법.

청구항 4

제1 항에 있어서,

상기 GCN 기반 위험도 예측 모델은 학습 데이터를 기반으로 생존 위험을 예측하도록 사전 학습된 이후 동결(fixed)되는 것을 특징으로 하는 병리 영상을 이용한 생존 시간 예측 방법.

청구항 5

제1 항 내지 제4 항 중 어느 하나의 항에 따른 방법을 수행하기 위한 프로그램 코드를 기록한 컴퓨터로 판독 가능한 기록 매체.

청구항 6

병리 영상을 전처리하여 조직(tissue) 영역을 추출한 후 각각 패치들로 잘라내는 전처리부;

상기 각각의 패치들을 그래프 구조로 구성하는 그래프 구성부; 및

상기 구성된 그래프 구조를 학습된 생존 시간 예측 모델에 적용하여 예측된 생존 시간을 출력하는 생존 시간 예측부를 포함하되,

상기 생존 시간 예측 모델은 GCN 기반 위험도 예측 모델과 GCN 기반 시간 예측 모델을 포함하되, 구성된 그래프 구조를 기학습된 GCN 기반 위험도 예측 모델에 적용하여 위험 특징 벡터를 추출하고, 상기 구성된 그래프 구조를 학습된 GCN 기반 시간 예측 모델에 적용하여 시간 특징 벡터를 추출한 후, 상기 위험 특징 벡터와 상기 시간 특징 벡터를 결합한 위험-시간 특징 벡터를 이용하여 예측된 생존 시간을 출력하는 것을 특징으로 하는 생존 시간 예측 장치.

청구항 7

제6 항에 있어서,

상기 그래프 구성부는,

상기 각각의 패치를 기학습된 종양 분류기에 적용하여 종양에 해당하는 패치들을 선별하고, 상기 선별된 각 패치들의 특징값을 각각 추출하고, 각 패치들의 특징값을 그래프의 꼭지점으로 각각 설정하며, 상기 각각의 패치 중심 위치 사이의 유클리드 거리를 기반으로 K-최근접 이웃 알고리즘을 적용하여 꼭지점 간의 에지를 각각 설정하여 상기 그래프 구조로 구성하는 것을 특징으로 하는 생존 시간 예측 장치.

발명의 설명**기술 분야**

[0001] 본 발명은 병리 영상을 이용하여 암 환자의 생존 시간을 예측하는 방법 및 그 장치에 관한 것이다.

배경 기술

[0003] 병리 영상(whole slide pathological image, WSI)에 대한 분석은 병리학자가 암 환자를 진단하고 예후를 예측하는 중요한 수단이다. 이러한 과정은 긴 시간과 노력을 필요로 하며, 병리학자의 숙련도에 따라 진단 결과가 달라지는 문제점이 있다.

[0004] 이에, 딥 러닝(deep learning) 기반 모델을 통해 자동으로 병리 영상을 분석하여 암 환자의 생존 예측을 수행하여 의사의 진단 오류를 줄이기 위한 다양한 연구들이 진행되고 있다. 그러나, 종래의 딥러닝 기반 생존 예측 모델(예를 들어, DeepAttnMISL, PatchGCN)은 환자의 사망 우선순위를 획득하기 위해 환자의 생존 위험을 예측하는 데만 중점을 두고 있는 단점이 있다.

발명의 내용**해결하려는 과제**

[0006] 본 발명은 병리 영상을 이용한 생존 시간 예측 방법 및 그 장치를 제공하기 위한 것이다.

[0007] 또한, 본 발명은 생존 위험도와 시간 예측을 통합하여 생존 시간 예측 정확도가 높은 병리 영상을 이용한 생존 시간 예측 방법 및 그 장치를 제공하기 위한 것이다.

[0008] 또한, 본 발명은 딥러닝 모델을 기반으로 정확한 생존 시간을 예측하면서도 환자 간의 사망 순서도 예측할 수 있어 의사의 진단 보조를 통해 환자의 삶의 질을 향상시킬 수 있는 병리 영상을 이용한 생존 시간 예측 방법 및 그 장치를 제공하기 위한 것이다.

과제의 해결 수단

[0010] 본 발명의 일 측면에 따르면 병리 영상을 이용한 생존 시간 예측 방법이 제공된다.

[0011] 본 발명의 일 실시예에 따르면, (a) 병리 영상을 전처리하여 조직(tissue) 영역을 각각 추출한 후 각각 패치들로 잘라내는 단계; (b) 상기 각각의 패치들을 그래프 구조로 구성하는 단계; (c) 상기 구성된 그래프 구조를 학습된 GCN 기반 위험도 예측 모델에 적용하여 위험 특징 벡터를 도출하는 단계; (d) 상기 구성된 그래프 구조를 학습된 GCN 기반 시간 예측 모델에 적용하여 시간 특징 벡터를 도출하는 단계; 및 (e) 상기 위험 특징 벡터와 상기 시간 특징 벡터를 결합한 위험-시간 특징 벡터를 이용하여 예측된 생존 시간을 출력하는 단계를 포함하는 병리 영상을 이용한 생존 시간 예측 방법이 제공될 수 있다.

[0012] 상기 (b) 단계는, 상기 각각의 패치를 기학습된 종양 분류기에 적용하여 종양에 해당하는 패치들을 선별하는 단계; 상기 선별된 각 패치들의 특징값을 각각 추출하고, 각 패치들의 특징값을 그래프의 꼭지점으로 각각 설정하는 단계; 및 상기 각각의 패치 중심 위치 사이의 유클리드 거리를 기반으로 K-최근접 이웃 알고리즘을 적용하여 꼭지점 간의 에지를 각각 설정하여 상기 그래프 구조로 구성하는 단계를 포함할 수 있다.

[0013] 상기 (e) 단계는, 상기 위험 특징 벡터의 순서 정보를 고려하여 상기 시간 특징 벡터가 보정되어 상기 위험-시간 특징 벡터가 생성될 수 있다.

[0014] 상기 GCN 기반 위험도 예측 모델은 학습 데이터를 기반으로 생존 위험을 예측하도록 사전 학습된 이후 동결(fixed)될 수 있다.

[0016] 본 발명의 다른 측면에 따르면, 병리 영상을 이용한 생존 시간 예측 장치가 제공된다.

[0017] 본 발명의 일 실시예에 따르면, 병리 영상을 전처리하여 조직(tissue) 영역을 추출한 후 각각 패치들로 잘라내는 전처리부; 상기 각각의 패치들을 그래프 구조로 구성하는 그래프 구성부; 및 상기 구성된 그래프 구조를 학습된 생존 시간 예측 모델에 적용하여 예측된 생존 시간을 출력하는 생존 시간 예측부를 포함하되, 상기 생존 시간 예측 모델은 GCN 기반 위험도 예측 모델과 GCN 기반 시간 예측 모델을 포함하되, 구성된 그래프 구조를 기학습된 GCN 기반 위험도 예측 모델에 적용하여 위험 특징 벡터를 추출하고, 상기 구성된 그래프 구조를 학습된 GCN 기반 시간 예측 모델에 적용하여 시간 특징 벡터를 추출한 후, 상기 위험 특징 벡터와 상기 시간 특징 벡터를 결합한 위험-시간 특징 벡터를 이용하여 예측된 생존 시간을 출력하는 것을 특징으로 하는 생존 시간 예측 장치가 제공될 수 있다.

[0018] 상기 그래프 구성부는, 상기 각각의 패치를 기학습된 종양 분류기에 적용하여 종양에 해당하는 패치들을 선별하고, 상기 선별된 각 패치들의 특징값을 각각 추출하고, 각 패치들의 특징값을 그래프의 꼭지점으로 각각 설정하며, 상기 각각의 패치 중심 위치 사이의 유클리드 거리를 기반으로 K-최근접 이웃 알고리즘을 적용하여 꼭지점 간의 에지를 각각 설정하여 상기 그래프 구조로 구성할 수 있다.

발명의 효과

[0020] 본 발명의 일 실시예에 따른 병리 영상을 이용한 생존 시간 예측 방법 및 그 장치를 제공함으로써, 생존 위험도와 시간 예측을 통합하여 생존 시간 예측 정확도를 높일 수 있는 이점이 있다.

[0021] 또한, 본 발명은 딥러닝 모델을 기반으로 정확한 생존 시간을 예측하면서도 환자 간의 사망 순서도 예측할 수 있어 의사의 진단 보조를 통해 환자의 삶의 질을 향상시킬 수 있는 이점도 있다.

도면의 간단한 설명

- [0023] 도 1은 본 발명의 일 실시예에 따른 병리 영상을 이용한 생존 시간 예측 방법을 나타낸 순서도.
 도 2는 본 발명의 일 실시예에 따른 그래프 구조 생성 과정을 설명하기 위해 도시한 도면.
 도 3은 본 발명의 일 실시예에 따른 GCN 기반 위험도 예측 모델의 구조를 도시한 도면.
 도 4는 본 발명의 일 실시예에 따른 생존 시간 예측 모델을 도시한 도면.
 도 5는 본 발명의 일 실시예에 따른 생존 시간 예측 모델에 대한 정량적 성능 비교 결과를 도시한 도면.
 도 6 및 도 7은 본 발명의 일 실시예에 따른 위험 특징 벡터 사용 여부에 대한 비교 실험 결과를 나타낸 도면.
 도 8은 본 발명의 일 실시예에 따른 생존 시간 예측 장치의 내부 구성을 개략적으로 도시한 블록도.

발명을 실시하기 위한 구체적인 내용

- [0024] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.
- [0025] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세히 설명한다.
- [0027] 도 1은 본 발명의 일 실시예에 따른 병리 영상을 이용한 생존 시간 예측 방법을 나타낸 순서도이고, 도 2는 본 발명의 일 실시예에 따른 그래프 구조 생성 과정을 설명하기 위해 도시한 도면이고, 도 3은 본 발명의 일 실시예에 따른 GCN 기반 위험도 예측 모델의 구조를 도시한 도면이며, 도 4는 본 발명의 일 실시예에 따른 생존 시간 예측 모델을 도시한 도면이고, 도 5는 본 발명의 일 실시예에 따른 생존 시간 예측 모델에 대한 정량적 성능 비교 결과를 도시한 도면이고, 도 6 및 도 7은 본 발명의 일 실시예에 따른 위험 특징 벡터 사용 여부에 대한 비교 실험 결과를 나타낸 도면이다.
- [0028] 단계 110에서 생존 시간 예측 장치(100)는 병리 영상(whole slide pathological image)에서 조직 영역(tissue region)을 추출한 후 작은 패치로 잘라낸다.
- [0029] 단계 115에서 생존 시간 예측 장치(100)는 패치들을 분류하여 종양(tumor)으로 분리된 패치들을 선별한다.
- [0030] 예를 들어, 생존 시간 예측 장치(100)는 도 2에 도시된 바와 같이, 병리 영상에서 각각 조직 영역을 추출한다. 이어, 생존 시간 예측 장치(100)는 각 조직 영역을 패치들로 각각 잘라낸다. 이어, 생존 시간 예측 장치(100)는 사전 학습된 종양 분류기(tumor classifier)에 패치들을 적용하여 종양에 해당하는 패치들을 선별한다. 여기서, 종양 분류기는 픽셀 레벨 종양 주석과 정상 주석이 있는 데이터 셋을 이용하여 사전 학습되어 있는 것을 가정하기로 한다.
- [0031] 단계 120에서 생존 시간 예측 장치(100)는 선별된 패치들을 이용하여 그래프(graph)를 구성한다.
- [0032] 이에 대해 보다 상세히 설명하기로 한다. 예를 들어, 병리 영상에서 배경 영역을 제거한다. 예를 들어, 'Otsu, N. A threshold selection method from gray-level histograms. IEEE Transactions on Syst. Man, Cybern. 9, 62-66'를 이용하여 배경 영역을 제거할 수도 있다. 배경 영역을 제거하는 방법은 반드시 이로 제한되는 것은 아니며 공지된 방법들은 제한 없이 적용될 수 있다.
- [0033] 생존 시간 예측 장치(100)는 전경 조직 영역을 512 x 512 크기의 겹치지 않는 패치들로 자른다.
- [0034] 종양 분류기는 ResNet-18과 같은 백본으로 구성되며, 픽셀 수준 관심 영역(ROI) 주석이 있는 병리 영상이 포함된 Camelyon17 데이터 셋을 사용하여 사전 학습될 수 있다.

- [0035] 종양 분류기에 의해 분리된 종양 패치들의 집합은 $\rho = \{\rho | f(\rho) > 0.5\}$ 과 같다. 여기서, f

는 입력 패치(ρ)의 중앙 확률을 나타낸다. 이어, 생존 시간 예측 장치(100)는 선별된 중앙 패치들이 서로 가까운 패치들과 연결되도록 그래프를 구성한다. 즉, 생존 시간 예측 장치(100)는 선별된 중앙 패치들의 특징을 ImageNet으로 기학습된 ResNet을 이용하여 추출한 후 각 특징을 그래프의 꼭지점으로 처리한다. 또한, 생존 시간 예측 장치(100)는 꼭지점 사이의 에지는 서로 가까운 패치가 상호 작용할 가능성이 더 높기 때문에 공간 거리 임계값(d)이 있는 패치 중심 위치 사이의 유클리드 거리를 사용하여 k-최근접 이웃 알고리즘으로 설정할 수 있다.

이를 수학적식으로 나타내면, 수학적식 1과 같이 나타낼 수 있다.

수학적식 1

$$e_{ij} = \begin{cases} 1, \text{if } j \in knn(v_i) \text{ and } D(v_i, v_j) \leq d, \\ 0, \text{otherwise,} \end{cases}$$

여기서, $e_{ij} \in E$ 는 두 정점(꼭지점) v_i, v_j 사이의 에지를 나타내고, $D(,)$ 는 두 패치의 중심 위치 사이의 유클리드 거리를 나타내며, $knn(v_i)$ 는 v_i 의 k-최근접 이웃을 나타낸다.

구성된 그래프는 $G = (V, E)$ 로 나타낼 수 있다. 여기서, V는 꼭지점 집합을 나타내고, E는 그래프의 에지의 집합을 나타낸다.

단계 125에서 생존 시간 예측 장치(100)는 구성된 그래프 표현을 GCN 기반 위험도 예측 모델에 적용하여 위험 특징 벡터를 추출한다.

GCN 기반 위험도 예측 모델은 도 3에 도시된 바와 같다. 도 3에 도시된 바와 같이, GCN 기반 위험도 예측 모델은 3개의 GCN 블록과 단일 GCN 계층을 포함하며, 그 뒤에 전역 최대 풀링과 전역 평균 풀링 및 2개의 완전 연결 계층(FC)를 포함한다.

GCN 블록은 수학적식 2와 같이 나타낼 수 있다.

수학적식 2

$$H^{(l+1)} = \sigma \left(\text{SagPool} \left(\mathcal{G} \left(H^{(l)} \right) \right) \right),$$

$$\mathcal{G} \left(H^{(l)} \right) = D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}} H^{(l)} W^{(l)},$$

여기서, $\mathcal{G}$ 는 하나의 GCN 계층에 대한 그래프 컨볼루션 연산을 나타내고, A는 인접 행렬이며, $\tilde{A} = A + I$ 는 자체 루프를 포함하는 인접 행렬을 나타내고, D는 대각 행렬이 $D_{ii} = \sum_j \tilde{A}_{ij}$ 와 같이 주어진 차수 행렬을 나타내며, $H^{(l)}$ 는 $H^{(0)} = V$ 인 1-번째 레이어에서 특징 행렬을 나타내고,

$W^{(l)}$ 는 레이어 l 을 위한 학습 파라미터를 나타내며, σ 는 활성화 함수를 나타내고, $SagPool$ 는 풀링 비율 r 을 가지는 셀프 어텐션 그래프 풀링을 나타낸다.

[0045] 그래프 컨볼루션 연산 이후, 글로벌 최대 풀링과 평균 풀링을 적용하여 위험 특징 벡터 h_r 이 획득될 수 있다.

GCN 기반 위험도 예측 모델의 학습이 완료된 후 분리되는 두개의 완전 연결 계층(FC)을 통해 특징 벡터(h_r)를 사용하여 생존 위험 R 을 예측하도록 학습될 수 있다. 이러한, GCN 기반 위험도 예측 모델은 Cox 모델을 사용하여 생존 위험(R)을 예측하도록 사전 훈련된 후 동결될 수 있다.

[0046] 평가 지표로 C-index가 사용되었다. C-index는 환자들 사이의 사망 순위가 얼마나 잘 구성되었는지를 평가하며, 수학적 3과 같이 정의될 수 있다.

수학적식 3

$$C_r = \frac{1}{K} \sum_{i=1}^P \delta_i \sum_{j: \tau_i < \tau_j} \mathbb{I}_{R_i > R_j},$$

[0047]

[0048] 여기서, C_r 는 위험 예측의 C-인덱스를 나타내고, K 는 비교 가능한 쌍의 전체 개수를 나타내며, P 는 총 환자

수를 나타내고, $\mathbb{I}_a$ 는 인디케이터 함수로, a 가 참이면 $\mathbb{I}_a = 1$ 이고, 그렇지 않으면

$$\mathbb{I}_a = 0$$

이다. 학습 단계에서 각 병리 영상은 개별 데이터 샘플로 간주될 수 있다. 검증 및 테스트 단계에서 한 환자에 대해 여러 병리 영상이 있는 경우 병리 영상 수준 예측의 평균 값을 환자 수준 예측 결과로 간주하였다.

[0049] 단계 130에서 생존 시간 예측 장치(100)는 구성된 그래프를 학습된 GCN 기반 시간 예측 모델에 적용하여 시간 특징 벡터를 도출한다. 여기서, GCN 기반 시간 예측 모델에 포함되는 GCN 구조는 GCN 기반 위험도 예측 모델의 구조와 동일하므로 중복되는 설명은 생략하기로 한다.

[0050] GCN 기반 시간 예측 모델 또한 학습 데이터를 기반으로 시간 특징 벡터를 출력하도록 사전 학습될 수 있다.

[0051] 단계 135에서 생존 시간 예측 장치(100)는 위험 특징 벡터(h_r)와 시간 특징 벡터(h_t)를 결합하여 위험-시간 특징 벡터(h_{rt})를 생성하고, 생성된 위험-시간 특징 벡터를 이용하여 환자의 생존 시간을 예측한다.

[0052] 본 발명의 일 실시예에 따르면, 생존 시간 예측 모델은 도 4에 도시된 바와 같이, GCN 기반 위험도 예측 모델 및 GCN 기반 시간 예측 모델, 복수의 FC 계층을 포함하여 구성된다. GCN 기반 위험도 예측 모델 및 GCN 기반 시간 예측 모델에는 동일한 그래프 구조(G)와 GCN 아키텍처가 사용된다.

[0053] 기 학습된 GCN 기반 위험도 예측 모델에 의해 추출된 위험 특징 벡터(h_r)은 GCN 기반 시간 예측 모델에서 추출된 시간 특징 벡터(h_t)와 연결되어 위험-시간 특징 벡터(h_{rt})가 도출될 수 있다. 위험 특징 벡터의 순서 정보를 고려하여 시간 특징 벡터가 보정되어 위험-시간 특징 벡터가 생성될 수 있다.

[0054] 서수 회귀 프레임워크(ordinal regression framework)에 따라 생존 시간 범위를 N 간격으로 나눈다. 예를 들어, $\{T_0, T_1, \dots, T_N\}$ 이 되, $T_0 < T_1 < \dots < T_N$ 이다. 간격은 학습 데이터베이스에서 환자의 생존 시간을 고르게 분배하기 위해 경험적으로 선택될 수 있다.

[0055] 각 환자의 이진 레이블을 수학적 4와 같이 정의하기로 한다.

수학적 4

$$b_n = \begin{cases} 1, & \text{if } \tau > T_n, \\ 0, & \text{otherwise,} \end{cases}$$

[0056]

[0057] 여기서, τ 은 생존 시간을 나타내고, n 은 시간 간격의 인덱스를 나타낸다.

[0058] 생존 시간 예측은 τ 가 T_n 보다 큰지 여부를 예측하는 n 번째 분류가 N 이진 분류를 해결하는 것을 목표로 하는 서수 회귀 문제로 변환될 수 있다.

[0059] GCN 기반 시간 예측 모델의 성능 평가를 위해 실제 데이터와 예상 생존 시간 사이의 평균 절대 오차(MAE)를 사용했다. 보다 상세하게, 검증 상태 δ , 예측된 생존 시간 $\tilde{\tau}$, 관찰 시간 τ 를 포함하는 각 환자에 대해 MAE는 수학적 5와 같이 계산될 수 있다.

수학적 5

$$\Delta = \frac{1}{P} \sum_{i=1}^P (\tau_i - \tilde{\tau}_i) \mathbb{I}_{\tilde{\tau}_i < \tau_i} + \delta_i (\tilde{\tau}_i - \tau_i) \mathbb{I}_{\tilde{\tau}_i \geq \tau_i}$$

[0060]

[0061] 또한, 수학적 6과 같이 C-인덱스를 계산하여 예측 생존 시간 간의 쌍별 순위 일관성을 평가했다.

수학적 6

$$C_t = \frac{1}{K} \sum_{i=1}^P \delta_i \sum_{j: \tau_i < \tau_j} \mathbb{I}_{\tilde{\tau}_i < \tilde{\tau}_j}$$

[0062]

[0063] 여기서, C_t 는 생존 시간 예측의 C-인덱스를 나타내고, GCN 기반 위험도 예측 모델의 성능 평가와 마찬가지로 학습 단계에서 각 병리 영상을 개인으로 간주하고 동일한 환자의 여러 병리 영상 결과를 평균하여 검증 및 테스트

트 단계에서 최종 예측 결과를 획득하였다.

[0064] 또한, 본 발명의 일 실시예에 따르면, 검증 인식 손실(censoring-aware loss)을 채택하여 사망이 관측된 데이터(검증되지 않은 데이터)와 사망이 관측되지 않은 데이터(검증된 데이터)를 모두 학습에 포함할 수 있다. 이러한, 손실 함수는 수학식 7과 같이 나타낼 수 있다.

수학식 7

$$L(\mathbf{o}, \mathbf{B}, \boldsymbol{\delta}) = - \sum_{n=1}^N [\delta_n + (1 - \delta_n)b_n] \times [b_n \log(o_n) + (1 - b_n) \log(1 - o_n)],$$

[0065] 여기서, $\mathbf{B} = \{b_1, b_2, \dots, b_N\}$ 이고, $\boldsymbol{\delta} = \{\delta_1, \delta_2, \dots, \delta_N\}$ 이다.

[0066] n -번째 분류의 출력은 o_n 으로 나타낼 수 있으며, $\mathbf{o} = \{o_1, o_2, \dots, o_N\}$ 이다.

[0067] 이는 사망이 관측된 검증되지 않은 트레이닝 샘플의 경우 교차 엔트로피가 모든 b_n 에 대해 계산됨을 의미한다. 그러나, 사망이 관측되지 않은 검증된 트레이닝 샘플의 경우, 마지막 관찰이 여전히 학습에 유용하다. 따라서, $b_n = 0$ 인 간격은 무시하고, $b_n = 1$ 인 경우에만 손실을 측정할 수 있다.

[0068] 최종 생존 예측 시간은 $\tilde{\tau}$ 은 수학식 8과 같이 계산될 수 있다.

수학식 8

$$\tilde{\tau} = \frac{T_{\sum_n \mathbb{I}_{o_n}} + T_{\sum_n \mathbb{I}_{o_n} + 1}}{2},$$

[0069] 여기서, $o_n \geq 0.5$ 이면 $\mathbb{I}_{o_n} = 1$ 이고, 그렇지 않으면 $\mathbb{I}_{o_n} = 0$ 이다.

[0070] 본 발명의 일 실시예에서는 타당성 검증을 위해 종래에 공개된 병리 영상 데이터 셋인 TCGA-BLCA, BRCA, GBM 데이터셋을 이용하여 실험을 수행하였다. 정량적 성능 비교 결과는 도 5에 도시된 바와 같다. 또한, 도 6 및 도 7은 본 발명의 일 실시예에 따른 위험 특징 벡터 사용 여부에 대한 비교 실험 결과를 나타낸 도면이다.

[0071] 도 5에서 보여지는 바와 같이, 3개의 데이터 셋 중에서 2개의 데이터 셋에서 종래 기술보다 더 정확한 생존 시간을 예측한 것을 알 수 있다.

[0072] 도 7의 (a)는 위험 특징 정보를 고려하지 않은 경우 예측된 생존 시간을 나타낸 도면이고, 도 7의 (b)는 위험 특징 정보를 고려한 경우 예측된 생존 시간을 나타낸 도면이다. 도 6 및 도 7에서 보여지는 바와 같이, 위험 특징 벡터와 시간 특징 벡터를 동시에 이용할 때 위험 특징 벡터가 순서 정보를 제공하여 가장 정확한 생존 시간을 예측할 수 있음을 알 수 있다.

- [0076] 도 8은 본 발명의 일 실시예에 따른 생존 시간 예측 장치의 내부 구성을 개략적으로 도시한 블록도이다.
- [0077] 도 8을 참조하면, 본 발명의 일 실시예에 따른 생존 시간 예측 장치(100)는 전처리부(810), 그래프 구성부(820), 생존 시간 예측부(830), 학습부(840), 메모리(850) 및 프로세서(860)를 포함하여 구성된다.
- [0078] 전처리부(810)는 병리 영상을 전처리하기 위한 수단이다.
- [0079] 예를 들어, 전처리부(810)는 병리 영상에서 배경 영역을 제거하고 조직 영역을 추출한 후 작은 패치들로 잘라낼 수 있다.
- [0080] 또한, 작은 패치들은 기학습된 종양 분류기에 의해 종양에 해당하는 패치들만 선별되며, 해당 종양에 해당하는 패치들만 이용하여 이후의 과정이 진행되는 것으로 이해되어야 할 것이다.
- [0081] 이는 도 1에서 전술한 바와 동일하므로 중복되는 설명은 생략하기로 한다.
- [0082] 그래프 구성부(820)는 선별된 종양 패치들을 이용하여 그래프를 구성한다.
- [0083] 예를 들어, 그래프 구성부(820)는 선별된 종양 패치들의 특징을 각각 추출한 후 종양 패치들의 각 특징을 꼭지점(정점)으로 설정하고, 꼭지점 사이의 에지는 공간 거리 임계값(d)이 있는 패치 중심 위치 사이의 유클리드 거리를 사용하여 k-최근접 이웃 알고리즘을 이용하여 설정할 수 있다.
- [0084] 이에 대해서도 도 1에서 설명한 바와 동일하므로 중복되는 설명은 생략하기로 한다.
- [0085] 생존 시간 예측부(830)는 생존 시간 예측 모델을 포함하며, 그래프 구조를 해당 생존 시간 예측 모델에 적용하여 환자의 생존 시간을 예측하기 위한 수단이다.
- [0086] 전술한 바와 같이, 생존 시간 예측 모델은 GCN 기반 위험도 예측 모델과 GCN 기반 시간 예측 모델을 포함하며, GCN 기반 위험도 예측 모델에서 예측된 위험 특징 벡터와 GCN 기반 시간 예측 모델에서 예측된 시간 특징 벡터를 결합하여 위험-시간 특징 벡터를 생성할 수 있다. 이어, 생존 시간 예측 모델은 해당 위험-시간 특징 벡터를 완전 연결 계층에 적용하여 생존 시간을 예측할 수 있다.
- [0087] 학습부(840)는 학습 데이터를 이용하여 생존 시간 예측 모델을 학습시키기 위한 수단이다.
- [0088] 메모리(850)는 본 발명의 일 실시예에 따른 병리 영상을 이용한 생존 시간 예측 방법을 수행하기 위한 명령어들을 저장한다.
- [0089] 프로세서(860)는 본 발명의 일 실시예에 따른 생존 시간 예측 장치의 내부 구성들(예를 들어, 전처리부(810), 그래프 구성부(820), 생존 시간 예측부(830), 학습부(840), 메모리(850) 등)을 제어하기 위한 수단이다.
- [0091] 본 발명의 실시 예에 따른 장치 및 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야 통상의 기술자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기매체(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.
- [0092] 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0093] 이제까지 본 발명에 대하여 그 실시 예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모

든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

부호의 설명

[0095]

100: 생존 시간 예측 장치

810: 전처리부

820: 그래프 구성부

830: 생존 시간 예측부

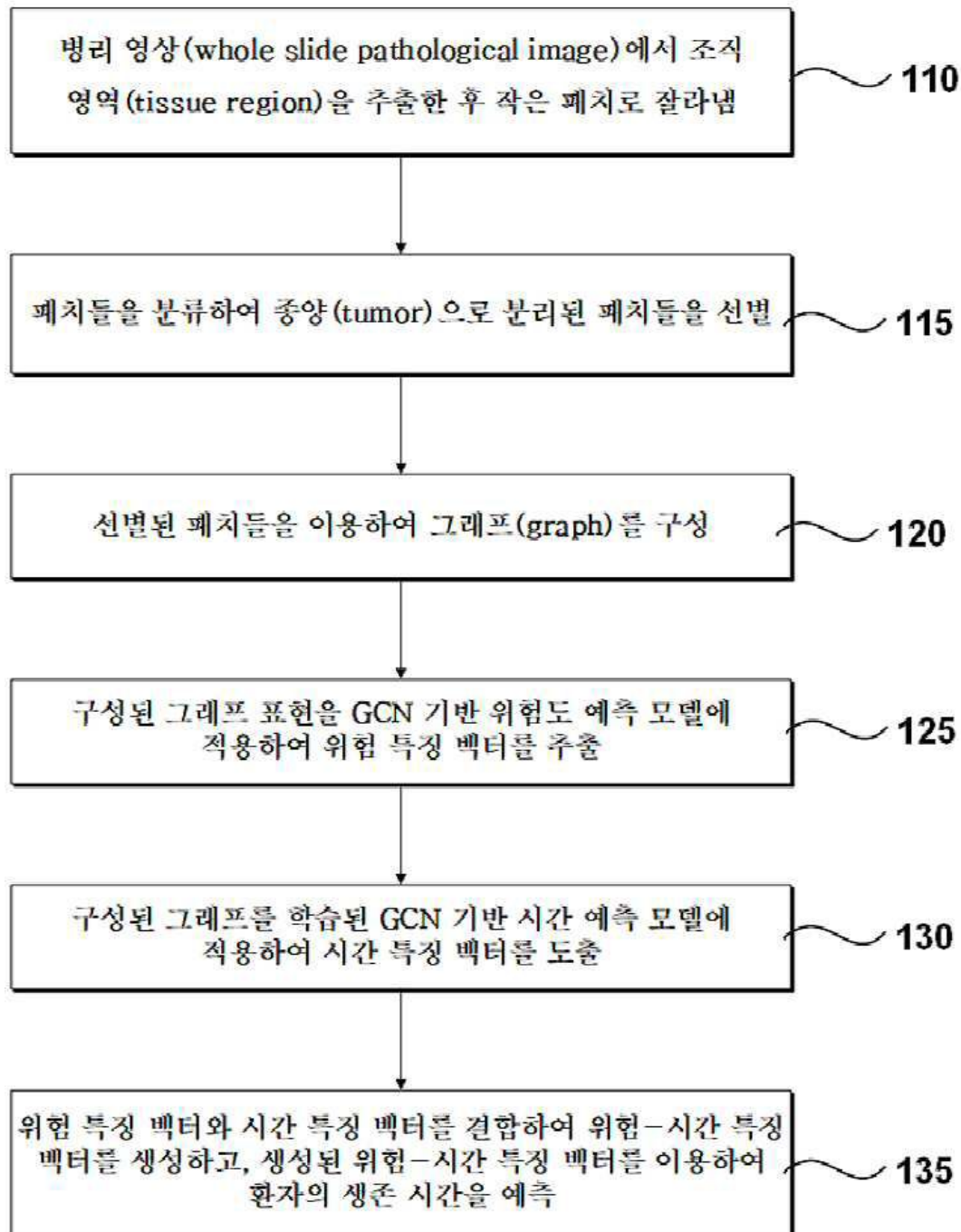
840: 학습부

850: 메모리

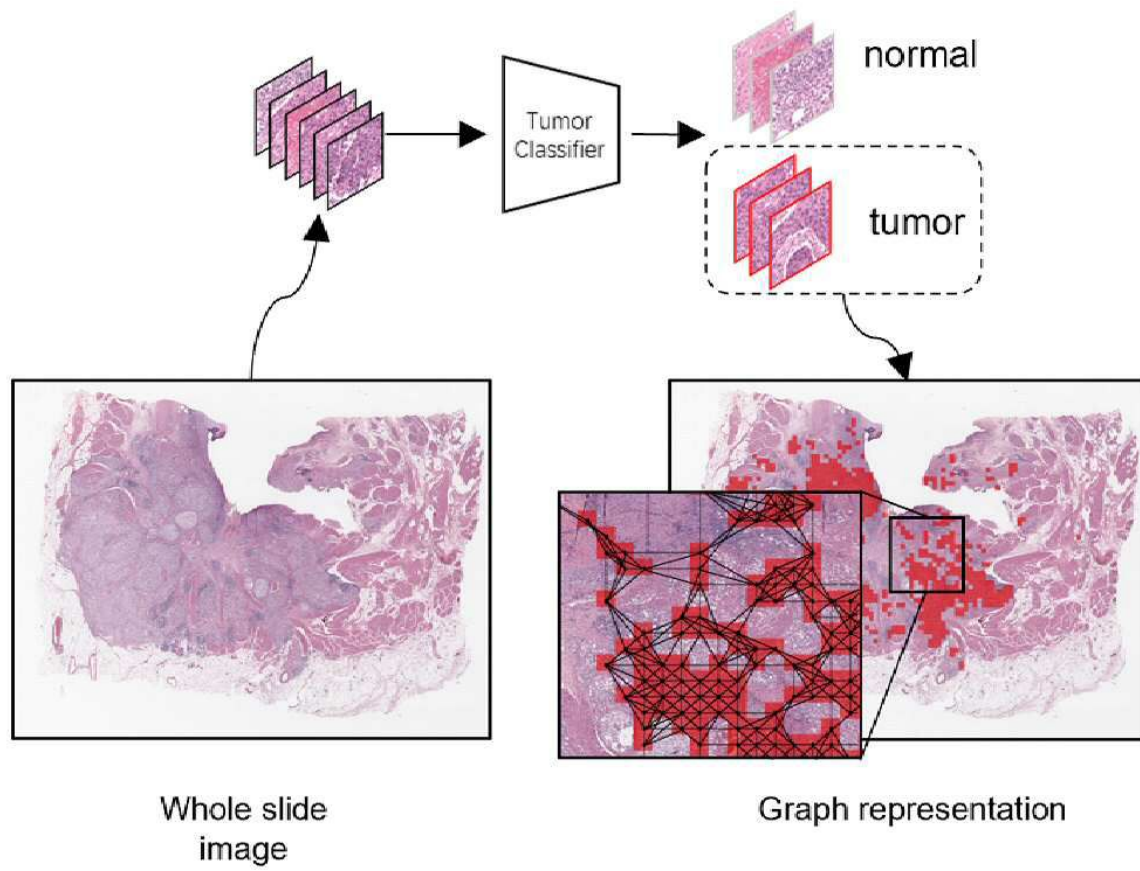
860: 프로세서

도면

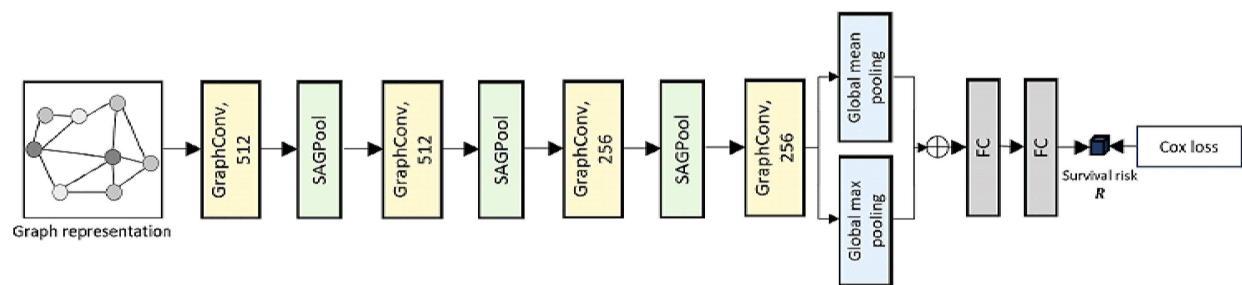
도면1



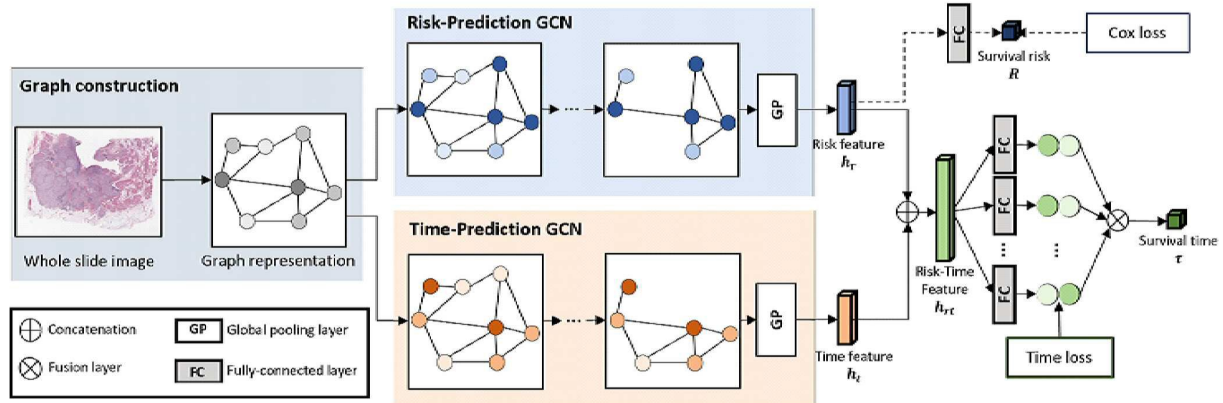
도면2



도면3



도면4



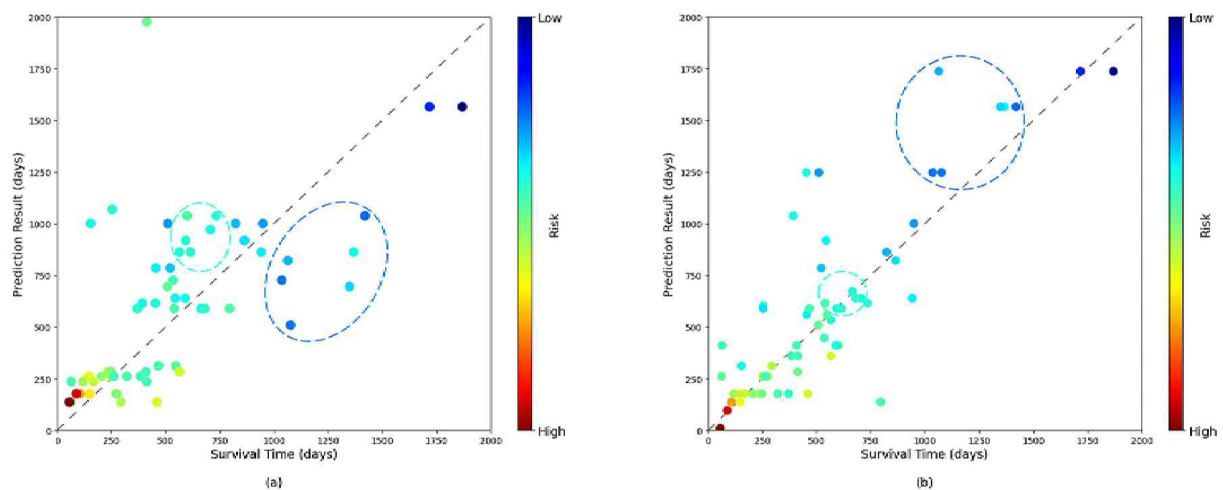
도면5

Method	BLCA				BRCA				GBM			
	MAE		C-index		MAE		C-index		MAE		C-index	
	Expect	Thresh	Risk	Time	Expect	Thresh	Risk	Time	Expect	Thresh	Risk	Time
DeepAttnMISL	386.1*	495.6*	0.526	-	172.6*	181.3*	0.652	-	310.0*	296.4*	0.626	-
PatchGCN	330.7*	304.5*	0.518	-	171.9*	205.1*	0.564	-	314.9*	300.5*	0.529	-
DeepGraphSurv	153.5*	167.1*	0.820	-	191.4*	193.9*	0.618	-	351.5*	337.8*	0.550	-
CDOR	329.6		-	0.550	203.3		-	0.561	320.0		-	0.492
Proposed	123.2		-	0.849	167.5		-	0.605	303.3		-	0.554

도면6

Method	BLCA			BRCA			GBM		
	MAE	C-index		MAE	C-index		MAE	C-index	
		Risk	Time		Risk	Time		Risk	Time
RP-GCN	-	0.834	-	-	0.627	-	-	0.563	-
TP-GCN	182.4	-	0.808	178.8	-	0.565	336.6	-	0.542
Proposed	123.2	-	0.849	167.5	-	0.605	303.3	-	0.554

도면7



도면8

