



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0115364
(43) 공개일자 2024년07월25일

(51) 국제특허분류(Int. Cl.)

A61B 5/00 (2021.01) G06T 7/11 (2017.01)
G16H 30/20 (2018.01) G16H 30/40 (2018.01)
G16H 50/20 (2018.01)

(52) CPC특허분류

A61B 5/4312 (2013.01)
A61B 5/004 (2013.01)

(21) 출원번호 10-2023-0007221

(22) 출원일자 2023년01월18일

심사청구일자 2023년01월18일

(71) 출원인

(주)제이엘케이

충청북도 청주시 청원구 오창읍 양청송대길 10,
204호(청주미래누리터(지식산업센터))

의료법인 성광의료재단

서울특별시 강남구 논현로 566 (역삼동)

(뒷면에 계속)

(72) 발명자

김동민

서울특별시 강남구 테헤란로64길14, 711호 (대치동, 더 나인 오피스텔)

이명재

서울특별시 강남구 학동로68길 29, 110동 902호
(삼성동, 삼성동힐스테이트1단지아파트)

(뒷면에 계속)

(74) 대리인

특허법인세원

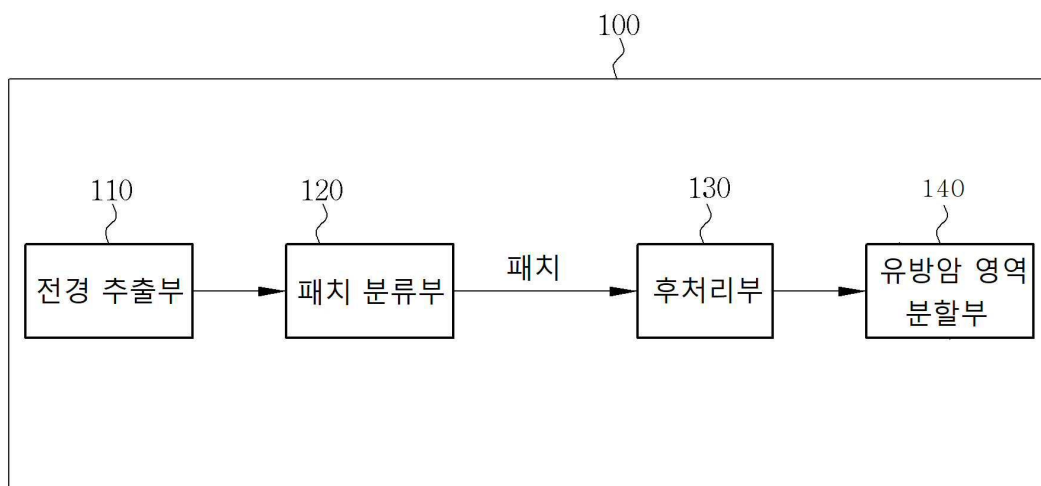
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 인공지능 기반 딥러닝 알고리즘을 이용한 유방암 영상의 종양 기질 비율 자동 산출 방법 및 시스템

(57) 요약

본 발명에 따른 인공지능 기반 딥러닝 알고리즘을 이용한 유방암 영상의 종양 기질 비율 자동 산출 시스템은 병리 슬라이드가 입력되면 상기 병리 슬라이드에서 배경을 제거하는 전경 추출부; 상기 병리 슬라이드로부터 이미지 패치들을 추출하고 상기 패치들 각각을 종양 패치 또는 정상 패치로 분류하는 패치 분류부; 각각의 패치에 대하여 인접 패치들의 모델 예측 결과들의 소정 확률이 종양 또는 정상을 나타내면, 해당 패치의 결과를 상기 소정 확률이 나타내는 종양 또는 정상으로 분류하는 후처리를 수행하는 후처리부; 및 종양으로 분류된 패치에 대해 침윤성 유방암 분할 과정을 수행하고 침윤성 종양 영역의 면적과 기질 영역의 면적의 비율을 이용해 종양-기질 비율을 산출하는 침윤성 유방암 영역 분할부를 포함한다.

대표도 - 도1



(52) CPC특허분류

G06T 7/11 (2017.01)
G16H 30/20 (2018.01)
G16H 30/40 (2018.01)
G16H 50/20 (2018.01)
G06T 2207/30068 (2013.01)

(71) 출원인

연세대학교 산학협력단

서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)

계명대학교 산학협력단

대구광역시 달서구 달구벌대로 1095, 계명대학교 산학협력관 201호(신당동)

(72) 발명자

강신욱

서울특별시 강남구 선릉로 221, 402동 1601호 (도곡동, 도곡렉슬아파트)

최홍국

경상남도 양산시 물금읍 야리로 89, 302동 902호(양산물금 양우 내안에 3차 센트럴 아파트)

문다민

서울특별시 강북구 솔샘로64길 15-17, 204호

이두희

서울특별시 성동구 독서당로 191, 7동 811호 (옥수동, 극동아파트)

조남훈

서울특별시 강남구 언주로130길 30, 103동 301호

박은향

서울특별시 서대문구 연세로 50-1, 의과대학 206호

박희정

서울특별시 서대문구 연세로 50-1, 의과대학 235호

신수진

서울특별시 강남구 언주로 211 강남세브란스병원

안희정

서울특별시 강남구 삼성로 14, 413동 1403호

강혜윤

서울특별시 성동구 매봉길 15, 108동 204호

김광일

서울특별시 강동구 고덕로 360, 338동 2604호

권선영

대구광역시 수성구 청호로 365, 303동 302호

정혜라

대구광역시 수성구 달구벌대로 2435, 108동 503호

이 발명을 지원한 국가연구개발사업

과제고유번호	1465034155
과제번호	HI21C0977040021
부처명	보건복지부
과제관리(전문)기관명	한국보건산업진흥원
연구사업명	디지털병리기반의암전문AI분석솔루션개발사업(R&D)
연구과제명	다중 인공지능 솔루션 연동 디지털 병리 플랫폼 개발
기 여 율	1/1
과제수행기관명	(주)제이엘케이
연구기간	2021.04.01 ~ 2025.12.31

명세서

청구범위

청구항 1

병리 슬라이드가 입력되면 상기 병리 슬라이드에서 배경을 제거하는 전경 추출부;

상기 병리 슬라이드로부터 이미지 패치들을 추출하고 상기 패치들 각각을 종양 패치 또는 정상 패치로 분류하는 패치 분류부;

각각의 패치에 대하여 인접 패치들의 모델 예측 결과들의 소정 확률이 종양 또는 정상을 나타내면, 해당 패치의 결과를 상기 소정 확률이 나타내는 종양 또는 정상으로 분류하는 후처리를 수행하는 후처리부; 및

종양으로 분류된 패치에 대해 침윤성 유방암 분할 과정을 수행하고 침윤성 종양 영역의 면적과 기질 영역의 면적의 비를 이용해 종양-기질 비율을 산출하는 침윤성 유방암 영역 분할부를 포함하는 것을 특징으로 하는 인공지능 기반 딥러닝 알고리즘을 이용한 유방암 영상의 종양 기질 비율 자동 산출 시스템.

청구항 2

청구항 1에 있어서,

상기 전경 추출부는 상기 병리 슬라이드에 대해 RGB(빨간색, 녹색, 파란색) 척도를 HSV(색조, 채도 및 값) 척도로 변환하고 상기 병리 슬라이드에 대해 채도 채널에서 채도 값이 4이상인 부분만 추출함으로써 상기 병리 슬라이드에서 배경을 삭제하는 것을 특징으로 하는 유방암 영상의 종양 기질 비율 자동 산출 시스템.

청구항 3

청구항 1에 있어서,

상기 패치 분류부는 종양 패치와 정상 패치 분류 모델을 학습하기 위해 ConvNeXt 모델을 사용하는 것을 특징으로 하는 유방암 영상의 종양 기질 비율 자동 산출 시스템.

청구항 4

청구항 1에 있어서,

상기 후처리부는 하나의 패치에 대하여 주변의 상하좌우 및 4개의 대각선에 해당하는 8개의 인접패치들을 고려하여 상기 후처리를 수행하는 것을 특징으로 하는 유방암 영상의 종양 기질 비율 자동 산출 시스템.

청구항 5

청구항 1에 있어서,

상기 소정 확률은 50% 보다 큰 것을 특징으로 하는 유방암 영상의 종양 기질 비율 자동 산출 시스템.

청구항 6

청구항 1에 있어서,

상기 후처리부는 상기 인접 패치들의 모델 예측 결과들이 동물이면 상기 해당 패치의 원래의 예측 결과를 사용한 것을 특징으로 하는 유방암 영상의 종양 기질 비율 자동 산출 시스템.

청구항 7

청구항 1에 있어서,

상기 침윤성 유방암 영역 분할부는 상기 병리 이미지에서 침윤성 유방암(Invasive cancer) 영역을 분할하기 위해 U-Net 기반 분할 모델을 사용한 것을 특징으로 하는 유방암 영상의 종양 기질 비율 자동 산출 시스템.

청구항 8

청구항 7에 있어서,

상기 침윤성 유방암 영역 분할부는 상기 U-Net 모델의 수축 경로에서 사용하는 네트워크를 InceptionResNet-v2로 변형시킨 모델을 사용한 것을 특징으로 하는 유방암 영상의 종양 기질 비율 자동 산출 시스템.

청구항 9

전경 추출부가 병리 슬라이드가 입력되면 상기 병리 슬라이드에서 배경을 제거하는 단계;

패치 분류부가 상기 병리 슬라이드로부터 이미지 패치들을 추출하고 상기 패치들 각각을 종양 패치 또는 정상 패치로 분류하는 단계;

후처리부가 각각의 패치에 대하여 인접 패치들의 모델 예측 결과들의 소정 확률이 종양 또는 정상을 나타내면, 해당 패치의 결과를 상기 소정 확률이 나타내는 종양 또는 정상으로 분류하는 단계; 및

침윤성 유방암 영역 분할부가 종양으로 분류된 패치에 대해 침윤성 유방암 분할 과정을 수행하고 침윤성 종양 영역의 면적과 기질 영역의 면적의 비율을 이용해 종양-기질 비율을 산출하는 단계를 포함하는 것을 특징으로 하는 인공지능 기반 딥러닝 알고리즘을 이용한 유방암 영상의 종양 기질 비율 자동 산출 방법.

발명의 설명

기술 분야

[0001] 본 발명은 인공지능 기반 딥러닝 알고리즘을 이용한 유방암 영상의 종양 기질 비율 자동 산출 방법 및 시스템에 관한 것이다.

배경 기술

[0003] 유방암은 여성에게 발생하는 암 중 가장 흔한 암이다. 보건복지부의 2019년 국가암등록사업 보고에 따르면 유방암 발생자수는 전체 여성암 발생자수의 20.6%를 차지한다. 유방암은 특히 발생률이 빠르게 증가하는 질환으로 2007년부터 2019년까지 유방암 발생률이 연평균 4.3%씩 증가하였다. 유방암의 발생률이 늘어나고 있음에 따라 유방암 진단 시 재발이나 생존과 관련된 예후 인자에 대한 연구가 더욱 중요해지고 있다.

[0004] 최근 디지털병리 분야에서는 예후 예측에 관한 연구가 활발하게 진행되고 있다. 예후와 관련성이 있는 인자 중 하나는 종양-기질 비율(Tumor Stroma Ratio, TSR)이다. 종양-기질 비율은 종양 내에 기질의 상대적 양을 나타낸다. 종양-기질 비율은 병리 전문의가 헤마톡실린 및 에오신(Hematoxylin and Eosin, H&E) 염색 조직 단면을 통해 측정한다[9]. 유방암 및 폐암 등 여러 종양에서 종양-기질 비율의 예후 관련성에 대한 많은 연구 결과가 있다. 종양 내에 기질의 비중이 높은 것은 나쁜 예후와 관련이 있다.

[0005] 하지만 종양-기질 비율은 병리 전문의가 스캔된 병리 슬라이드를 보고 육안으로 판단하므로 의사들 간 불일치율이 높다. 선행 연구에 따르면, 관찰자 간 종양-기질 비율 측정 상관관계는 0.239와 0.486 사이였다 (Cohen's kappa). 최근 인공지능을 사용한 정량적인 분석을 통해 의사들 간 시각적 평가의 불일치 문제를 개선하는 연구가 진행되고 있다[15]. 그러나 인공지능을 활용하여 병리 슬라이드에서 종양-기질 비율을 자동으로 산출하는 알고리즘 개발 연구는 거의 없는 현실이다.

[0006] 최근 디지털병리 인공지능 연구는 국소적인 분석에 국한된 패치 단위 학습의 한계를 극복하고자 하는 시도가 활발하다. 병리 의사는 진단 시 저배율로 조직의 전반적인 특징들(Global features)을 확인한 후 고배율로 국소적인 특징들(Local features)을 확인한다. 이 경우, 중앙 영역의 구분을 위해 병리 슬라이드를 패치화하였는데, 패치화는 넓은 범위를 볼 수 없다는 단점이 있다.

[0007] 서울대학교 연구진은 그래프 딥러닝 방식을 사용하여 병리 영상 내의 세포 간 상호작용을 나타내는 연구결과를 발표하였다. 그들은 병리 영상을 그래프로 표현하여 세포 간 상호작용의 해석이 가능하도록 하여 패치 단위 학습법의 한계를 극복하였다.

발명의 내용

해결하려는 과제

[0009] 본 발명은 전술한 문제를 해결하기 위해 안출된 것으로서, 인접패치를 고려한 간단한 후처리 방식을 적용하여 주변 패치와의 상호작용을 고려함으로써 패치 단위 학습의 한계를 극복하고자 한 인공지능 기반 딥러닝 알고리즘을 이용한 유방암 영상의 중앙 기질 비율 자동 산출 방법 및 시스템을 제공한다.

과제의 해결 수단

[0011] 전술한 문제를 해결하기 위한 본 발명의 일실시예에 따른 인공지능 기반 딥러닝 알고리즘을 이용한 유방암 영상의 중앙 기질 비율 자동 산출 시스템은 병리 슬라이드가 입력되면 상기 병리 슬라이드에서 배경을 제거하는 전경 추출부; 상기 병리 슬라이드로부터 이미지 패치들을 추출하고 상기 패치들 각각을 중앙 패치 또는 정상 패치로 분류하는 패치 분류부; 각각의 패치에 대하여 인접 패치들의 모델 예측 결과들의 소정 확률이 중앙 또는 정상을 나타내면, 해당 패치의 결과를 상기 소정 확률이 나타내는 중앙 또는 정상으로 분류하는 후처리부를 수행하는 후처리부; 및 중앙으로 분류된 패치에 대해 침윤성 유방암 분할 과정을 수행하고 침윤성 중앙 영역의 면적과 기질 영역의 면적의 비를 이용해 중앙-기질 비율을 산출하는 침윤성 유방암 영역 분할부를 포함한다.

[0012] 상기 전경 추출부는 상기 병리 슬라이드에 대해 RGB(빨간색, 녹색, 파란색) 척도를 HSV(색조, 채도 및 값) 척도로 변환하고 상기 병리 슬라이드에 대해 채도 채널에서 채도 값이 4이상인 부분만 추출함으로써 상기 병리 슬라이드에서 배경을 삭제할 수 있다.

[0013] 상기 패치 분류부는 중앙 패치와 정상 패치 분류 모델을 학습하기 위해 ConvNeXt 모델을 사용할 수 있다.

[0014] 상기 후처리부는 하나의 패치에 대하여 주변의 상하좌우 및 4개의 대각선에 해당하는 8개의 인접패치들을 고려하여 상기 후처리를 수행할 수 있다.

[0015] 상기 소정 확률은 50% 보다 클 수 있다.

[0016] 상기 후처리부는 상기 인접 패치들의 모델 예측 결과들이 동일이면 상기 해당 패치의 원래의 예측 결과를 사용할 수 있다.

[0017] 상기 침윤성 유방암 영역 분할부는 상기 병리 이미지에서 침윤성 유방암(Invasive cancer) 영역을 분할하기 위해 U-Net 기반 분할 모델을 사용할 수 있다.

[0018] 침윤성 유방암 영역 분할부가 상기 U-Net 모델의 수축 경로에서 사용하는 네트워크를 InceptionResNet-v2로 변형시킨 모델을 사용할 수 있다.

[0019] 또한, 본 발명의 일실시예에 따른 인공지능 기반 딥러닝 알고리즘을 이용한 유방암 영상의 중앙 기질 비율 자동 산출 방법은 전경 추출부가 병리 슬라이드가 입력되면 상기 병리 슬라이드에서 배경을 제거하는 단계; 패치 분류부가 상기 병리 슬라이드로부터 이미지 패치들을 추출하고 상기 패치들 각각을 중앙 패치 또는 정상 패치로 분류하는 단계; 후처리부가 각각의 패치에 대하여 인접 패치들의 모델 예측 결과들의 소정 확률이 중앙 또는 정상을 나타내면, 해당 패치의 결과를 상기 소정 확률이 나타내는 중앙 또는 정상으로 분류하는 단계; 및 침윤성 유방암 영역 분할부가 중앙으로 분류된 패치에 대해 침윤성 유방암 분할 과정을 수행하고 침윤성 중앙 영역의 면적과 기질 영역의 면적의 비를 이용해 중앙-기질 비율을 산출하는 단계를 포함한다.

발명의 효과

[0022] 본 발명에서는 인공지능을 활용하여 유방암 병리 영상에서 종양-기질 비율을 자동으로 산출하는 알고리즘을 개발하였다. 종양-기질 비율 계산을 위해 총 2개의 인공지능 모델을 사용하였다. 제1 인공지능 모델은 각 패치(Patch)를 종양과 정상으로 분류하는 모델이다. 제2 인공지능 모델은 제1 인공지능 모델에서 종양으로 분류된 패치에 대해 침윤성 유방암 영역을 분할하는 모델이다. 현재는 병리 의사가 병리 슬라이드를 육안으로 보고 종양-기질 비율을 판독하므로 관측자의 주관이 개입될 수 있다. 그에 반해 본 발명에서 제시한 인공지능 기반 종양-기질 비율을 자동 산출하는 방식은 관측자의 주관이 개입되지 않는 객관적이고 정량적인 방법이다. 인공지능을 사용하여 종양-기질 비율을 자동으로 산출한다면 병리 의사들 간의 불일치성을 줄이고 의사들의 업무 부담을 줄일 수 있을 것으로 기대된다.

[0023] 본 발명을 기반으로 삼아 앞으로 국내 여러 병원들에서 수집한 유방암 병리 슬라이드와 임상 데이터를 사용하여 인공지능 기반으로 산출된 종양-기질 비율을 이용한 예후 예측 연구를 진행할 예정이다. 또한 전이학습을 적용하여 유방암 외 타 암종 병리 슬라이드에 적용 가능한 종양-기질 비율 산출 모델을 개발할 예정이다.

도면의 간단한 설명

[0025] 도 1은 본 발명의 일실시예에 따른 인공지능 기반 딥러닝 알고리즘을 이용한 유방암 영상의 종양 기질 비율 자동 산출 시스템의 구성을 도시한 도면이다.

도 2는 도 1의 유방암 영상의 종양 기질 비율 자동 산출 시스템에서 입력되는 전체 슬라이드 이미지를 처리한 결과들의 예들을 나타낸 도면이다.

도 3은 ConvNeXt 모델의 구조를 나타낸 도면이다.

도 4는 각 모델의 성능 평가 결과는 나타낸 도면이다.

도 5는 병리의사 어노테이션 예시이다.

도 6은 본 발명에 따른 하나의 패치와 인접패치들을 나타낸 도면이다.

도 7은 후처리 전후의 각각이 패치들의 예측 결과를 나타낸 도면이다.

도 8은 전체 슬라이드 이미지에 대한 종양 및 정상 패치 분류 및 후처리 결과를 도시한다.

도 9는 InceptionResNet-v2 모델의 구성을 나타낸 도면이다.

도 10은 InceptionResNet-v2 네트워크를 위한 스키마(Schema)를 나타낸 도면이다.

도 11은 원본 이미지 패치 및 침윤성 유방암과 기질을 나타낸 도면이다.

도 12는 종양-기질 비율을 산출한 결과를 나타낸 도면이다.

도 13은 본 발명의 일실시예에 따른 클러스터링 알고리즘 기반의 효율적인 어노테이션 검수 작업 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0026] 본 발명은 다양한 변환을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변환, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0027] 다만, 실시형태를 설명함에 있어서, 관련된 공지 기능 혹은 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그에 대한 상세한 설명은 생략한다. 또한, 도면에서의 각 구성요소들의 크기는 설명을 위하여 과장될 수 있으며, 실제로 적용되는 크기를 의미하는 것은 아니다.

- [0028] 또한, 명세서 전체에서, 일 구성요소가 다른 구성요소와 "연결된다" 거나 "접속된다" 등으로 언급된 때에는, 상기 일 구성요소가 상기 다른 구성요소와 직접 연결되거나 또는 직접 접속될 수도 있지만, 특별히 반대되는 기재가 존재하지 않는 이상, 중간에 또 다른 구성요소를 매개하여 연결되거나 또는 접속될 수도 있다고 이해되어야 할 것이다. 또한, 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.
- [0029] 도 1은 본 발명의 일실시예에 따른 인공지능 기반 딥러닝 알고리즘을 이용한 유방암 영상의 종양 기질 비율 자동 산출 시스템의 구성을 도시한 도면이다. 도 2는 도 1의 유방암 영상의 종양 기질 비율 자동 산출 시스템에서 입력되는 전체 슬라이드 이미지를 처리한 결과들의 예들을 나타낸 도면이다.
- [0030] 이하 도 1 내지 도 2를 참조하여 유방암 영상의 종양 기질 비율 자동 산출 시스템을 설명하기로 한다.
- [0031] 본 발명의 일실시예에 따른 유방암 영상의 종양 기질 비율 자동 산출 시스템(100)은 컴퓨터 단말, 서버 또는 전용 장치로 구성되거나, 각 기능을 제공하는 구성요소가 각각 컴퓨터 단말, 서버 또는 전용 장치로 구성될 수 있다. 또는, 본 발명의 일실시예에 따른 유방암 영상의 종양 기질 비율 자동 산출 시스템(100)은 각 기능을 제공하는 각각의 구성요소가 하드웨어 또는 소프트웨어로 구성될 수 있다.
- [0032] 보다 구체적으로, 본 발명의 일실시예에 따른 유방암 영상의 종양 기질 비율 자동 산출 시스템(100)은 전경 추출부(110), 패치 분류부(120), 후처리부(130) 및 침윤성 유방암 영역 분할부(140)를 포함한다.
- [0033] 전경 추출부(110)는 병리 슬라이드가 입력되면 이 슬라이드에서 배경을 제거하는 전경 추출(Foreground extraction) 과정을 수행한다. 즉, 전경 추출부(110)는 전체 슬라이드 이미지(Whole Slide Image)에서 조직 영역을 자동으로 추출하기 위해 전경 추출 기술을 사용한다. 종래 소개되었던 전경 추출 연구들은 다음과 같다. Bug 등은 필터 기반 알고리즘을 개발했으며, Bandi 등은 조직 영역 분할을 위해 컨볼루션 딥러닝 방식으로 접근하였다. 기존 방법들과 달리, 본 발명에서는 보다 더 단순한 전경 추출 방법을 사용하였다.
- [0034] 구체적으로, 전경 추출부(110)는 각 슬라이드에 대해 RGB(빨간색, 녹색, 파란색) 척도를 HSV(Hue, Saturation, Value)(색조, 채도 및 값) 척도로 변환한다. 색조, 채도 및 값(H, S, V)은 0과 255 사이의 값으로 표현된다. 채도값 S는 특정한 색상의 가장 진한 상태를 100% 하였을 때 진함의 정도를 나타낸다. 채도값 0은 같은 명도의 무채색을 나타낸다. 전경 추출부(110)는 각 슬라이드에 대해 S (채도) 채널에서 S 값이 4이상인 부분만 추출한다. 또한, 노이즈 감소를 위해 확장과 침식연산을 적용할 수 있다.
- [0035] 전경 추출부(110)는 병리 슬라이드에서 배경을 제거한 후 병리 슬라이드를 패치 분류부(120)에 제공한다.
- [0036] 다시 말해, 패치 분류부(120)는 병리 슬라이드로부터 이미지 패치들을 추출한 후 각 이미지 패치를 종양과 정상 패치로 분류하고 종양과 정상 패치 분류 결과를 후처리부(130)에 제공한다.
- [0037] 종양 패치 또는 정상 패치의 분류를 위해 패치 분류부(120)는 각 패치를 종양과 정상으로 분류하는 인공지능 모델을 사용할 수 있다.
- [0038] 상기 인공지능 모델은 종양 영역의 구분을 위한 모델로서, 패치 단위 분류 인공지능 모델을 학습하였다. 학습을 위하여 어노테이션된 종양과 정상의 각 영역에서 512×512 픽셀($256 \times 256 \mu\text{m}$) 크기의 이미지 패치를 추출하였다.
- [0039] 구체적으로, 병리 영상은 svf 파일의 형식으로 취득할 수 있으며, 각 이미지의 크기는 한 번의 길이가 100,000 픽셀 이상으로 매우 크다. 병리 영상들의 크기가 다양하고 또한 크기(1 GB 초과)가 매우 크기 때문에 인공 신경망을 통해 직접 처리하기에는 적합하지 않다. 인공 신경망으로 병리 영상을 처리하기 위해 이미지를 더 작은 크기(예, 512×512 픽셀)로 패치화 하여 학습하는 방법이 널리 연구되고 있다. 병리 영상의 인공지능 학습을 위한 패치화를 위하여 Openslide 라이브러리를 사용할 수 있다.
- [0040] 일반적으로 병리 이미지 딥러닝 연구에서 256×256 픽셀 또는 512×512 픽셀 크기의 패치를 많이 사용한다. 예컨대, 종양 80,000 장, 정상 21,035 장의 패치가 생성되었으며, 두 클래스 간 균형을 맞추어 학습하기 위하여 종양 80,000 장의 패치 중 21,035 장의 패치를 랜덤으로 선별하여 학습에 사용하였다. 모델 성능 테스트를 위하여, 종양 패치 2,890 장과 정상 패치 2,733 장은 테스트 셋으로 남겨두고 나머지 데이터들을 사용하여 학습을 진행하였다. ConvNeXt, VGG16, ResNet50 모델을 각각 사용하여 종양과 정상 패치 분류 모델을 학습할 수 있다.
- [0041] 테스트 데이터셋에서 측정한 ConvNeXt 모델의 종양과 정상 패치 분류 성능을 표 1에 표기하였다. ResNet50, VGG16 모델과 비교하였을 때 ConvNeXt 모델의 분류 성능이 더 뛰어남을 확인할 수 있었다. 수신자조작특성

(Receiver Operating Characteristic, ROC) 곡선을 통해 각 모델의 성능 평가 결과는 도 4에 나타내었다.

표 1

	AUROC	Accuracy	Specificity	Sensitivity
ConvNeXt[22]	0.963	0.925	0.977	0.875
ResNet50[19]	0.930	0.899	0.959	0.842
VGG16[37]	0.956	0.910	0.947	0.876

[0042]

[0043]

일 실시예에 따라 본 발명에서는 종양 패치와 정상 패치 분류 모델을 학습하기 위해 ConvNeXt 모델을 사용하였다.

[0044]

ConvNeXt 모델은 Zhuang Liu 등이 제안하였다. ConvNeXt는 비전 트랜스포머(Vision Transformers)의 설계에서 영감을 받은 컨볼루션 모델(ConvNet)이며, ImageNet top-1 accuracy 87.8%의 성능을 달성하였다.

[0045]

ConvNeXt 모델의 구조는 도 3에 나타나 있다. $d7 \times 7$ 은 7×7 커널 사이즈를 사용한 깊이합성곱(Depthwise convolution)을 의미한다. 깊이합성곱을 사용해서 수용필드(Receptive field)를 넓히고 연산량을 줄일 수 있다. 또한 정규화 계층(Normalization layer(LN))을 사용하였고, 활성화함수(Activation function)는 GELU(Gaussian Error Linear Unit)를 사용하였다.

[0046]

종양 영역에서 추출된 패치는 종양 패치 그리고 정상 영역에서 추출된 패치는 정상 패치로 라벨링(Labeling)하였다. 그리고, QuPath 소프트웨어를 사용하여 정상(Normal), 종양(Tumor), 침윤성 유방암 영역을 어노테이션하였다.

[0047]

도 5는 병리의사 어노테이션 예시이다. 정상 영역은 노랑, 종양 영역은 빨강, 침윤성 유방암 영역은 파란색으로 표시하였다.

[0048]

본 발명에서 사용한 ConvNeXt 기반 종양-정상 분류 모델과 U-Net 기반 침윤성 유방암 분할 모델은 파이토치(PyTorch) 라이브러리를 이용하여 구현할 수 있다. 그래픽 카드는 엔비디아(NVIDIA) RTX A6000을 사용하였다.

[0049]

후처리부(130)는 종양과 정상 패치 분류 결과에 대해 8개의 인접패치를 고려한 후처리(Post-processing) 방식을 적용한다.

[0050]

일반적으로, 병리영상을 작은 크기의 패치로 나누어서 학습하는 패치 단위 학습 방법은 주변 패치들 간의 관계를 고려하지 않는다는 한계가 있다. 이러한 한계점을 개선하기 위해서, 본 발명에서는 주변 패치들의 예측 결과를 고려하는 후처리 기법을 통해 종양 영역 구분 성능을 향상시키고자 하였다.

[0051]

도 6은 본 발명에 따른 하나의 패치와 인접패치들을 나타낸 도면이다.

[0052]

도 6을 참조하면, 하나의 패치 (P_{ij})에 대하여 주변 상하좌우, 4곳의 대각선에 해당하는 8개의 인접패치를 고려하였다.

[0053]

후처리부(130)는 각각의 패치에 대하여 인접 패치들의 모델 예측 결과의 과반수가 종양이면 해당 패치의 결과를 종양으로 하는 후처리를 수행한다.

[0054]

후처리부(130)는 각각의 패치에 대하여 인접 패치들의 모델 예측 결과들의 소정 확률이 종양 또는 정상을 나타내면, 해당 패치의 결과를 상기 소정 확률이 나타내는 종양 또는 정상으로 분류한다.

[0055]

다시 말해, 후처리부(130)는 인접 패치들의 모델 예측 결과의 과반수가 정상이면 해당 패치의 결과를 정상으로 하는 후처리를 수행한다. 예컨대, 후처리부(130)가 각각의 패치에 대하여 인접 패치들의 모델 예측 결과들의 50% 보다 큰 확률로 종양 또는 정상을 나타내면, 해당 패치의 결과를 종양이나 정상으로 하는 후처리를 수행한다.

[0056]

이 경우, 후처리부(130)는 인접 패치들의 모델 예측 결과가 동률이면 원래의 예측 결과를 사용한다.

[0057]

도 7은 후처리 전후의 각각의 패치들의 예측 결과를 나타낸 도면이다.

[0058] 도 7(a)는 후처리 이전 각각의 패치들에 대한 인공지능 모델 예측 결과의 예시를 도시한다. 중앙에 위치한 패치에 대해 8개의 인접 패치를 고려한 결과, 종양으로 예측된 패치 7개, 정상으로 예측된 패치가 1개로 종양으로 예측된 패치가 더 많음을 알 수 있다. 따라서, 후처리 과정에서 중앙에 위치한 패치의 예측결과를 정상에서 종양으로 수정할 수 있다. 도 7(b)는 도 7(a)의 인공지능 모델 예측 결과를 보정한 예시를 도시한다. 본 발명에서는 후처리를 진행하기 전과 후의 종양 그리고 정상 패치 분류 정확도를 비교하고자 하였다.

[0059] 상기 후처리 방식을 적용한 결과 테스트 데이터셋 분류 정확도가 92.46%에서 93.24%로 향상되었다. 전체 슬라이드 이미지에 패치 분류 모델과 후처리를 적용한 결과는 표 2 및 도 8에 있다. 인접 패치를 고려한 후처리 적용 결과, 후처리를 적용하기 전보다 종양 영역을 정교하게 구분함을 확인할 수 있다.

표 2

		Predicted			
		ConvNeXt		Post-Processing	
		Normal	Tumor	Normal	Tumor
Actual	Normal	2671	62	2712	21
	Tumor	362	2528	359	2531
Accuracy[%]		92.46		93.24	

[0060]

[0061] 도 8을 참조하면, 전체 슬라이드 이미지에 대한 종양 및 정상 패치 분류 및 후처리 결과를 도시하는데, 도 8(a)는 병리학자 어노테이션을 나타내며, 도 8(b)는 종양 및 정상 분류 모델 예측 결과를 나타내면 도 8(c) 후처리 결과를 나타낸다.

[0062] 도 8(c)에 도시된 바와 같이, 본 발명의 후처리를 수행함으로써, 종양 및 정상 영역이 병리학자의 어노테이션과 일치함을 알 수 있다. 즉, 본 발명에 따라 후처리를 수행함으로써, 종양 그리고 정상 패치 분류 정확도가 향상됨을 알 수 있다.

[0063] 후처리부(130)는 후처리를 적용한 종양 영역 분류 결과를 침윤성 유방암 영역 분할부(140)에 제공한다.

[0064] 침윤성 유방암 영역 분할부(140)는 종양으로 분류된 패치에 대해 침윤성 유방암 분할 과정을 수행하고 이 과정을 통해 얻은 침윤성 유방암 영역의 면적과 기질 영역의 면적의 비를 이용해 종양-기질 비율을 산출한다. 전술한 바와 같이, 예후와 관련성이 있는 인자 중 하나는 종양-기질 비율(Tumor Stroma Ratio, TSR)이다. 종양-기질 비율은 종양 내에 기질의 상대적 양을 나타낸다.

[0065] 구체적으로, 침윤성 유방암 영역 분할부(140)는 종양-기질 비율의 자동 산출을 위한 인공지능 모델로서 침윤성 유방암 영역의 분할을 위한 모델을 사용한다.

[0066] 침윤성 유방암 영역 분할부(140)는 침윤성 유방암 영역의 분할을 위한 모델을 사용하여, 종양 영역 분류 결과에서 종양-기질 비율을 산출할 수 있다.

[0067] 일 실시예에 따라, 본 발명에서는 병리 이미지에서 침윤성 유방암(Invasive cancer) 영역을 분할하기 위해 U-Net 기반 분할 모델을 사용할 수 있다.

[0068] U-Net은 의생명(Biomedical) 분야에서 이미지 분할(Image Segmentation)을 목적으로 제안된 합성곱 네트워크(Convolutional Network) 기반 모델이다. U-Net은 Olaf Ronneberger 등이 제안하였으며, U-Net을 기반으로 한 딥러닝 모델은 의료 이미지 분할에 우수한 성능을 보임이 알려져 있다. U-Net 모델의 구조는 도 7에 설명되어 있다. U-Net 모델은 수축 경로(Contracting path)와 확장 경로(Expansion path)로 구성되어 있다. 수축 경로는 모델 구조의 중심을 기준으로 좌측에 해당하는 부분으로 다운 샘플링(Downsampling) 과정이다. 다운 샘플링 시마다 특징 지도(Feature map)의 크기는 1/2배, 필터(Filter) 수는 2배가 된다. 다운 샘플링을 반복적으로 실행하여 다양한 깊이에서 이미지의 의미정보(Context information)를 추출한다. 확장 경로는 모델 구조의 우측에 해당하는 부분으로 업 샘플링(Upsampling) 과정이다. 업 샘플링 시마다 특징 지도의 크기는 2배, 필터의 수는 1/2배가 된다. 업 샘플링을 반복적으로 실행하는 과정에서 수축 경로의 대칭되는 계층에 연결(Concatenation)함으로써 수축 경로의 각각의 계층에서 얻은 의미 정보들을 기반으로 다시 원래의 의미지로 확장해 나가는 과정

을 갖는다. 마지막 단계로, 특징 지도에 1x1 필터 합성곱(Convolution)을 거쳐 최종 출력을 얻는다.

[0069] 또한, 본 발명에서는 U-Net 모델의 수축 경로에서 사용하는 네트워크를 InceptionResNet-v2로 변형시킨 모델을 사용하였다. InceptionResNet-v2 모델은 residual connection과 Inception 구조를 결합시킨 모델이다.

[0070] 도 9는 InceptionResNet-v2 모델의 구성을 나타낸 도면이고 도 10은 InceptionResNet-v2 네트워크를 위한 스키마(Schema)를 나타낸 도면이다.

[0071] 도 9 및 도 10을 참조하면, InceptionResNet-v2 모델은 Stem 블록, Inception 블록, reduction 블록, 평균 풀링(Average pooling) 층, dropout 층, 소프트맥스 함수로 구성되어 있다. Stem 블록은 6개의 컨볼루션 블록과 1개의 최대 풀링(Max pooling) 층으로 구성되어 있고, Inception-resnet-A 블록은 7개의 컨볼루션 블록으로 구성된 인셉션 모듈 5개로 구성되어 있고, Inception-resnet-B 블록은 5개의 컨볼루션 블록으로 구성된 인셉션 모듈 10개로 구성되어 있고, Inception-resnet-C 블록은 5개의 컨볼루션 블록으로 구성된 인셉션 모듈 5개로 구성되어 있다.

[0072] 병리 이미지에서 침윤성 유방암(Invasive cancer) 영역을 분할하기 위한 U-Net 기반 분할 모델은 전체 패치들 중에 침윤성 유방암 어노테이션 영역이 패치 전체 면적의 20% 이상인 패치들을 학습에 사용하였다. 총 22,079 쌍의 이미지와 정답지 패치 쌍이 생성되었으며 모델 성능 테스트를 위하여, 2,432 쌍의 패치들을 테스트 셋으로 남겨두고 나머지 데이터들을 사용하여 학습을 진행하였다.

[0073] 침윤성 유방암 영역을 분할하기 위하여 U-Net 모델의 수축 경로의 네트워크를 InceptionResNet-v2로 교체한 InceptionResNet-v2 backboneed U-Net 모델을 사용하였다. 모델의 성능은 수학적 1 내지 수학적 3의 Dice coefficient, 재현율(Recall), 정밀도(Precision)의 3개의 지표로 측정하였다.

수학적 1

[0074]

$$\text{Dice} = \frac{2TP}{2TP + FP + FN}$$

수학적 2

[0075]

$$\text{Recall} = \frac{TP}{TP + FN}$$

수학적 3

[0076]

$$\text{Precision} = \frac{TP}{TP + FP}$$

[0077] 상기 수학적 1 내지 수학적 3에서 TP, TN, FP 및 FN은 각각 true positive, true negative, false positive 및 false negative를 의미한다. 표 3은 각각의 정의를 설명하는 표이다.

표 3

		Ground Truth	
		Cancer(255)	Background(0)
Prediction	Cancer(255)	TP	FP
	Background(0)	FN	TN

[0078]

[0079]

침윤성 유방암 분할 정답지와 인공지능 예측 이미지에서 침윤성 유방암 영역에 해당하는 픽셀의 화소 강도 값은 255이고 배경에 해당하는 픽셀의 화소 강도 값은 0이다. 정답지와 예측 이미지의 화소 강도 값을 비교하였을 때, TP는 두 이미지 모두 255인 픽셀의 개수, TN은 두 이미지 모두 0인 픽셀의 개수, FP는 정답지는 0, 예측 이미지는 255인 픽셀의 개수, FN은 정답지는 255, 예측 이미지는 0인 픽셀의 개수이다. 도 11(a)는 원본 이미지 패치이며, 도 11(b)의 흰색 영역은 침윤성 유방암 그리고 검정색 영역은 기질에 해당한다.

[0080]

테스트 병리 슬라이드에서 종양과 정상 분류 모델과 침윤성 유방암 분할 모델을 적용하였고 종양-기질 비율을 산출한 결과를 도 12에 나타내었다.

[0081]

예컨대, 침윤성 유방암 영역 분할부(140)는 종양 영역의 면적과 기질 영역의 면적을 해당 영역의 픽셀 개수를 계수하여 산출할 수 있다. 도 12에 도시된 바와 같이, 종양-기질 비율이 0.5 이상이면 Stroma-High, 0.5 미만이면 Stroma-Low로 분류하였다.

[0082]

테스트 데이터셋에서 측정한 U-Net 모델과 FPN 모델의 침윤성 유방암 분할 성능은 표 4에 표기하였다. 두 모델 모두 수축 경로에서 사용하는 네트워크를 InceptionResNet-v2로 변형시켜 학습에 사용하였다. Dice coefficient 성능 지표를 이용하여 두 모델의 분할 성능을 비교한 결과, U-Net 모델의 침윤성 유방암 분할 성능이 더 우수함을 확인하였다. 테스트 데이터 셋에서 측정한 U-Net 모델의 Dice coefficient 성능 지표는 0.741이었다. 다소 낮은 수치를 보이는 것은 어노테이션을 수작업으로 생성하기 때문에 생기는 어노테이션의 부정확성으로 인한 것이다. 병리 분야에서 정확한 분할 어노테이션을 얻는 것은 어렵다. 수치로 나타낼 수는 없지만, U-Net 모델은 침윤성 유방암 영역을 잘 분할하였다.

표 4

	Dice coefficient	Precision	Recall
U-Net[16]	0.741	0.799	0.756
FPN[40]	0.726	0.816	0.721

[0083]

[0084]

도 13은 본 발명의 일실시예에 따른 클러스터링 알고리즘 기반의 효율적인 어노테이션 검수 작업 방법을 설명하기 위한 흐름도이다.

[0085]

도 13은 본 발명의 일실시예에 따른 인공지능 기반 딥러닝 알고리즘을 이용한 유방암 영상의 종양 기질 비율 자동 산출 방법의 흐름도이다.

[0086]

먼저, 전경 추출부(110)가 병리 슬라이드가 입력되면 상기 병리 슬라이드에서 배경을 제거한다(S210).

[0087]

이어서, 패치 분류부(120)가 상기 병리 슬라이드로부터 이미지 패치들을 추출하고(S220), 상기 패치들 각각을 종양 패치 또는 정상 패치로 분류한다(S230).

[0088]

후처리부(130)가 종양과 정상 패치 분류 결과에 대해 8개의 인접 패치를 고려한 후처리를 수행한다(S240). 구체적으로 후처리부(130)는 종양 패치 또는 정상 패치로 분류된 각 패치에 대해 인접 패치들의 모델 예측 결과들의 소정 확률이 종양 또는 정상을 나타내면, 해당 패치의 결과를 상기 소정 확률이 나타내는 종양 또는 정상으로 분류한다(S240). 상기 소정 확률은 51% 이상인 것이 바람직하다.

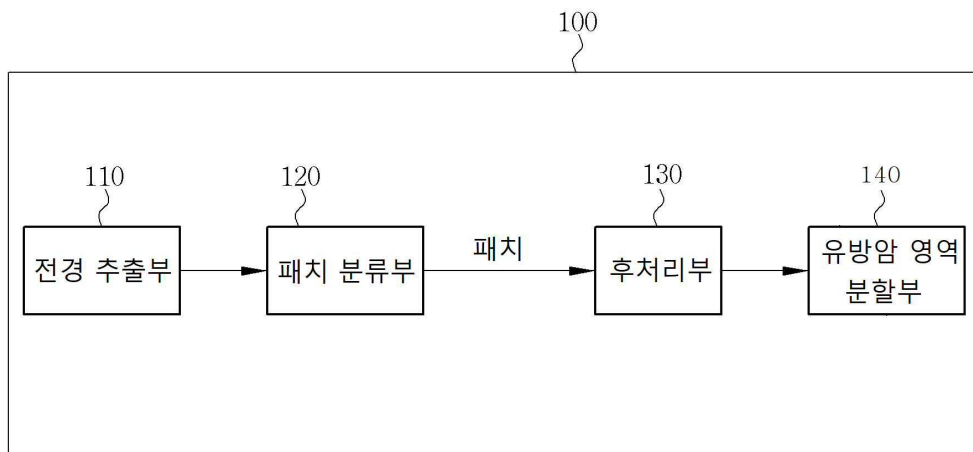
- [0089] 침윤성 유방암 영역 분할부(140)가 종양으로 분류된 패치에 대해 침윤성 유방암 분할 과정을 수행하고(S250), 침윤성 종양 영역의 면적과 기질 영역의 면적의 비를 이용해 종양-기질 비율을 산출한다(S260).
- [0090] 전술한 바와 같은 본 발명의 상세한 설명에서는 구체적인 실시예에 관해 설명하였다. 그러나 본 발명의 범주에서 벗어나지 않는 한도 내에서는 여러 가지 변형이 가능하다. 본 발명의 기술적 사상은 본 발명의 전술한 실시예에 국한되어 정해져서는 안 되며, 청구범위뿐만 아니라 이 청구범위와 균등한 것들에 의해 정해져야 한다.

부호의 설명

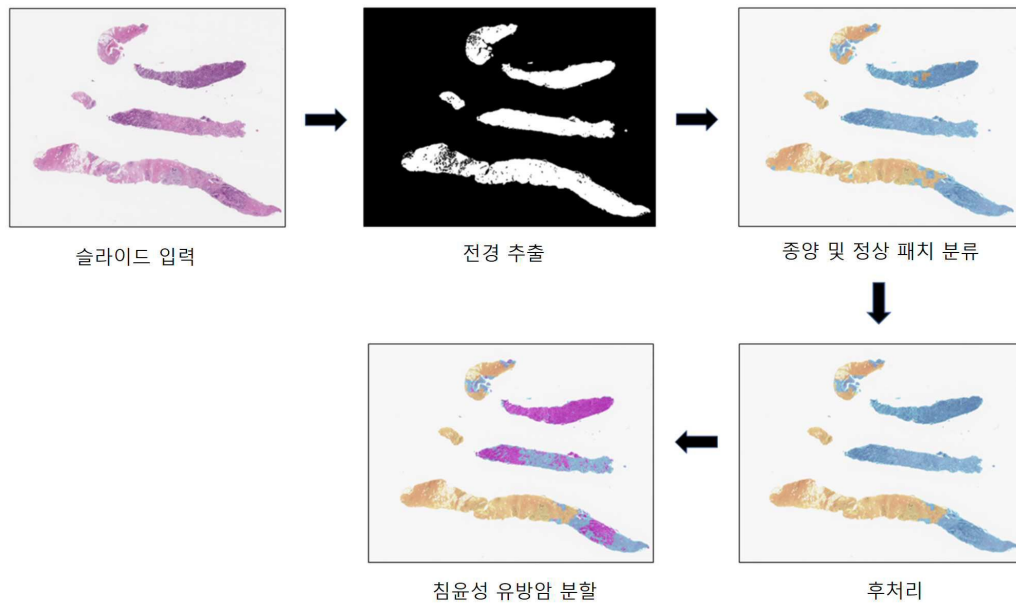
- [0092] 100: 유방암 영상의 종양 기질 비율 자동 산출 시스템
- 110: 전경 추출부
- 120: 패치 분류부
- 130: 후처리부
- 140: 유방암 영역 분할부

도면

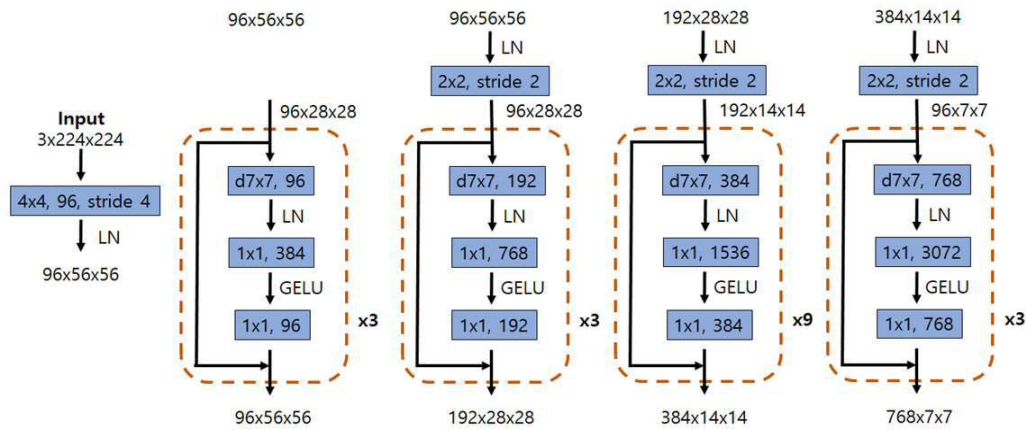
도면1



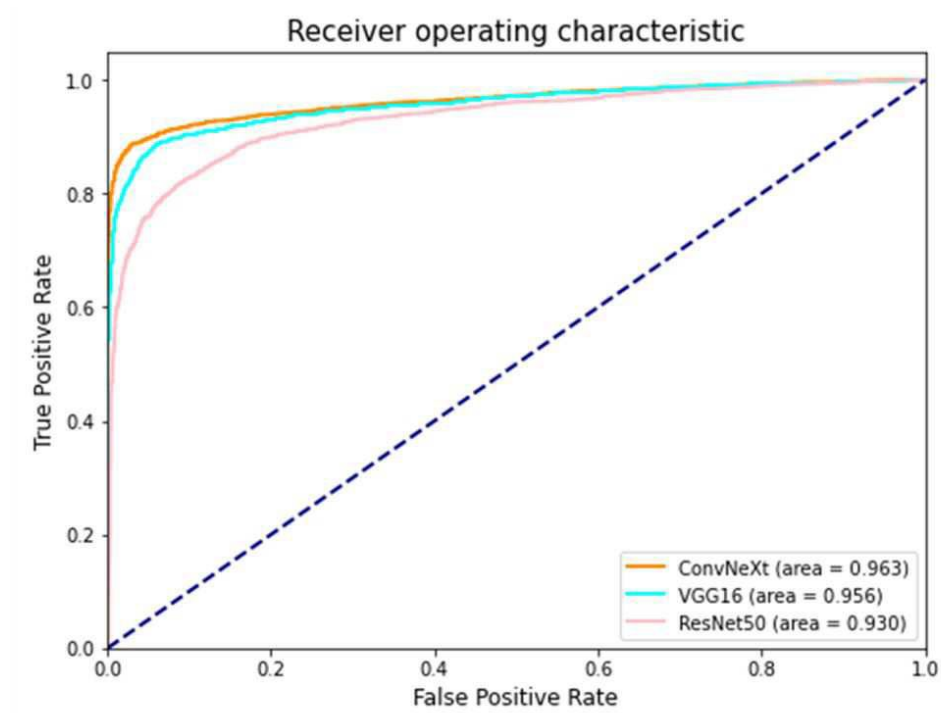
도면2



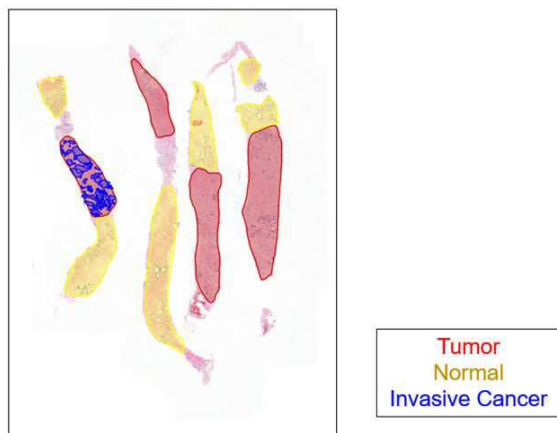
도면3



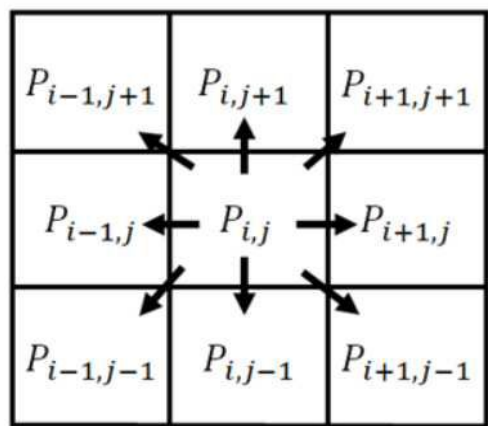
도면4



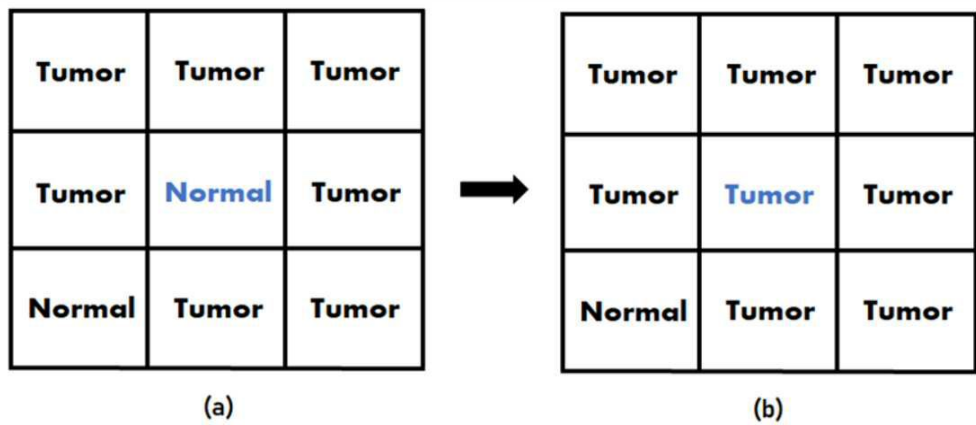
도면5



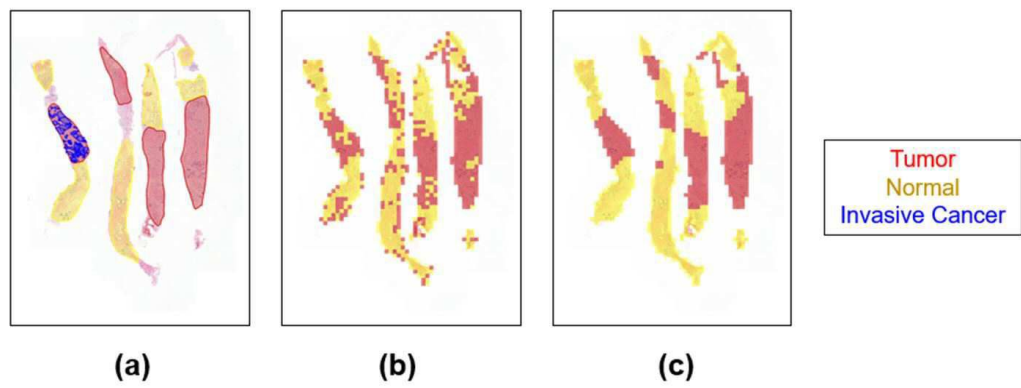
도면6



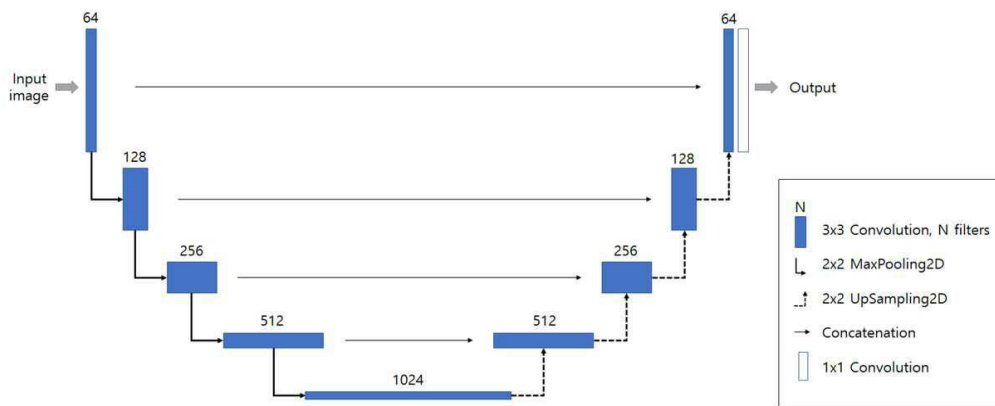
도면7



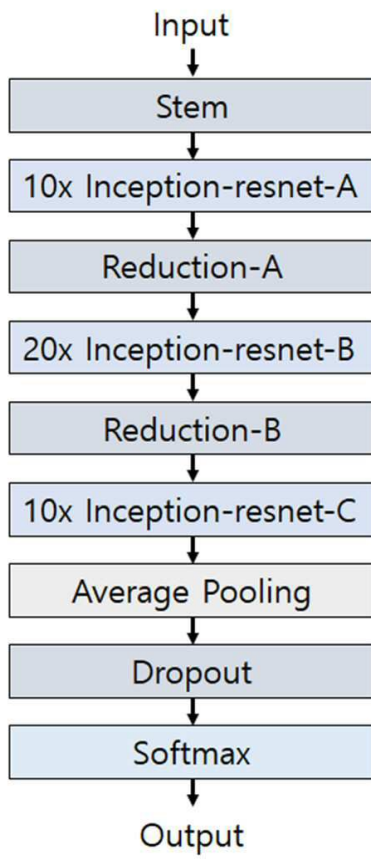
도면8



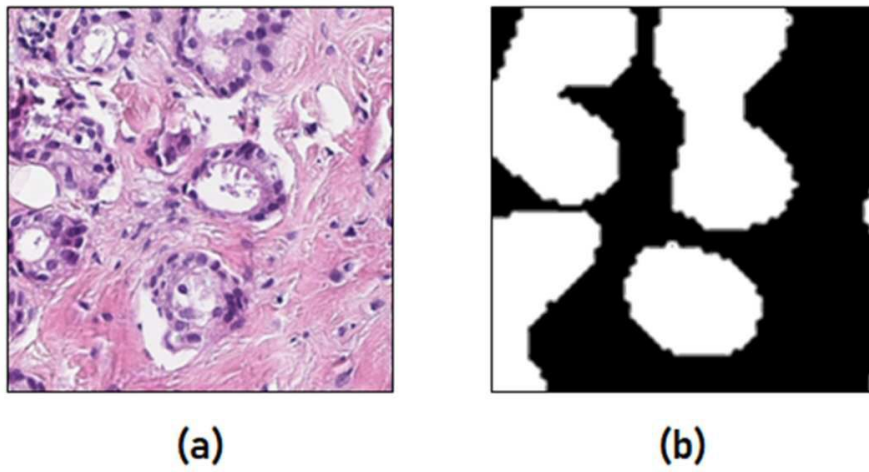
도면9



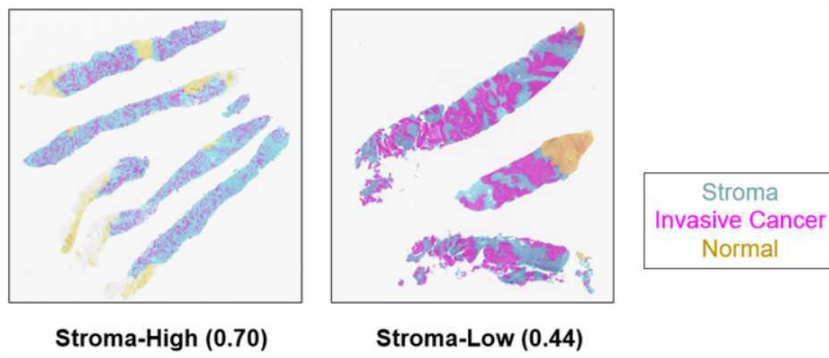
도면10



도면11



도면12



도면13

