



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0034752
(43) 공개일자 2025년03월11일

(51) 국제특허분류(Int. Cl.)

G16H 50/50 (2018.01) A61B 6/03 (2025.01)
G06F 16/33 (2025.01) G06F 40/279 (2020.01)
G16H 10/60 (2018.01) G16H 15/00 (2018.01)
G16H 30/20 (2018.01) G16H 50/20 (2018.01)

(52) CPC특허분류

G16H 50/50 (2018.01)
A61B 6/032 (2013.01)

(21) 출원번호 10-2023-0117095

(22) 출원일자 2023년09월04일

심사청구일자 2023년09월04일

(71) 출원인

연세대학교 산학협력단

서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

이충근

서울특별시 성동구 금호로 107, 101동 603호 (금호동2가, 래미안하이리버)

김승섭

서울특별시 마포구 서강로 95, 103동 1102호 (창전동, 창전동삼성아파트)

(뒷면에 계속)

(74) 대리인

유미특허법인

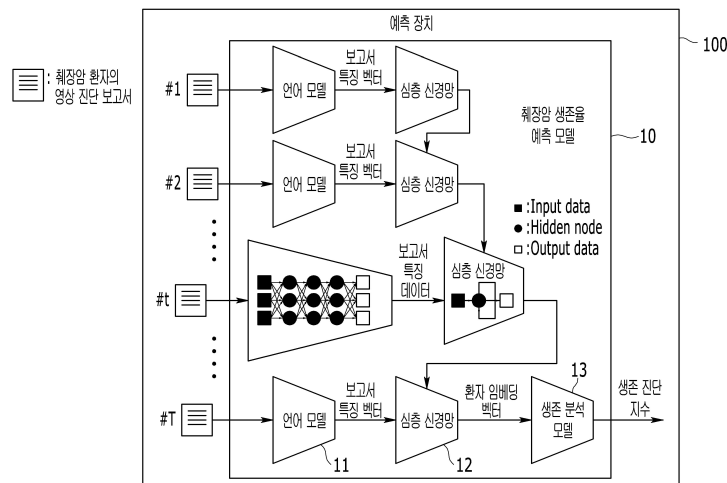
전체 청구항 수 : 총 22 항

(54) 발명의 명칭 영상 진단 보고서를 이용한 cheated 환자의 생존율 예측 방법, 예측 장치, 훈련 방법 및 훈련 장치

(57) 요약

적어도 하나의 프로세서에 의해 동작하는 예측 장치의 예측 방법으로서, cheated 환자에 대한 의사의 임상적 소견이 기재된 영상 진단 보고서를 입력 받는 단계, 그리고 입력 데이터로부터 cheated 생존 가능성인 cheated 생존율을 예측하도록 훈련된 cheated 생존율 예측 모델을 이용하여, 상기 입력 받은 영상 진단 보고서로부터 상기 cheated 환자의 생존율을 예측하는 단계를 포함한다.

대표도



(52) CPC특허분류

G06F 16/3344 (2019.01)

G06F 16/3347 (2019.01)

G06F 40/279 (2020.01)

G06N 3/044 (2023.01)

G16H 10/60 (2021.08)

G16H 15/00 (2021.08)

G16H 30/20 (2018.01)

G16H 50/20 (2018.01)

(72) 발명자

강재우

서울특별시 성동구 왕십리로 241, 103동 1402호 (
행당동, 서울숲 더샵)

김선규

인천광역시 연수구 앵고개로104번길 22, 705동 30
1호(동춘동, 송도파크자이)

명세서

청구범위

청구항 1

적어도 하나의 프로세서에 의해 동작하는 예측 장치의 예측 방법으로서,
 췌장암 환자에 대한 의사의 임상적 소견이 기재된 영상 진단 보고서를 입력 받는 단계, 그리고
 입력 데이터로부터 췌장암 생존 가능성인 췌장암 생존율을 예측하도록 훈련된 췌장암 생존율 예측 모델을 이용하여, 상기 입력 받은 영상 진단 보고서로부터 상기 췌장암 환자의 생존율을 예측하는 단계
 를 포함하는, 예측 방법.

청구항 2

제1항에서,
 상기 예측하는 단계는,
 상기 영상 진단 보고서를 자연어 처리 모델에 입력하여 상기 영상 진단 보고서에 대한 복수개의 단어 특징 벡터를 추출하고, 추출한 복수개의 단어 특징 벡터들을 통합한 단일 특징 벡터를 보고서 특징 벡터로 추출하며, 추출한 보고서 특징 벡터를 생존 분석 모델에 입력하여 상기 췌장암 생존율을 획득하는, 예측 방법.

청구항 3

제2항에서,
 상기 자연어 처리 모델은,
 ClinicalBERT(Bidirectional Encoder Representations from Transformers) 모델인, 예측 방법.

청구항 4

제2항에서,
 상기 예측하는 단계는,
 상기 자연어 처리 모델을 통해 연속된 복수개의 영상 진단 보고서로부터 추출한 보고서 특징 벡터를 순환 신경망(Recurrent Neural Network, RNN)을 통해 연속적으로 분석하여 환자 임베딩 벡터를 추출하고, 추출한 환자 임베딩 벡터를 상기 생존 분석 모델에 입력하여 상기 췌장암 생존율을 획득하는, 예측 방법.

청구항 5

제2항에서,
 상기 생존 분석 모델은,
 콕스 비례 위험 모델(Cox proportional-hazards)을 포함하고,
 상기 췌장암 생존율은,
 콕스 비례 위험도(Cox Proportional-Hazard Rate)인, 예측 방법.

청구항 6

적어도 하나의 프로세서에 의해 동작하는 예측 장치로서,

입력 데이터로부터 채장암 생존 가능성인 채장암 생존율을 예측하도록 훈련된 인공지능 모델로서, 입력 받은 채장암 환자에 대한 의사의 임상적 소견이 기재된 영상 진단 보고서로부터 상기 채장암 환자의 생존율을 예측하는 채장암 생존율 예측 모델

을 포함하는, 예측 장치.

청구항 7

제6항에서,

상기 채장암 생존율 예측 모델은,

상기 영상 진단 보고서에 대한 복수개의 단어 특징 벡터를 통합한 단일 특징 벡터를 보고서 특징 벡터로 추출하는 자연어 처리 모델, 그리고

상기 보고서 특징 벡터를 입력 받아 상기 채장암 생존율을 출력하는 생존 분석 모델

을 포함하는, 예측 장치.

청구항 8

제7항에서,

상기 영상 진단 보고서는,

채장암 환자에 대한 일정 기간 동안의 컴퓨터 단층 촬영(computed tomography scan, CT scan)에 대한 방사선학 보고서인, 예측 장치.

청구항 9

제7항에서,

상기 채장암 생존율 예측 모델은,

상기 자연어 처리 모델에 의해 추출된 연속된 복수개의 영상 진단 보고서로부터 각각의 보고서 특징 벡터를 순차적으로 입력 받고, 입력 받은 보고서 특징 벡터에 대한 연속 분석을 통해 환자 임베딩 벡터를 추출하여 상기 생존 분석 모델에 입력하는 순환 신경망(Recurrent Neural Network, RNN)을 더 포함하는, 예측 장치.

청구항 10

제9항에서,

상기 순환 신경망은,

GRU(gated rectified unit)를 사용하여 보고서 특징 벡터를 연속적으로 분석해서 환자 임베딩 벡터를 출력하는, 예측 장치.

청구항 11

적어도 하나의 프로세서에 의해 동작하는 훈련 장치로서,

채장암 환자의 영상 진단 보고서를 학습 데이터로 사용하여, 적어도 하나의 태스크를 학습한 채장암 생존 예측

모델을 생성하는 모델 학습부를 포함하고,

상기 췌장암 생존 예측 모델은,

췌장암 환자에 대한 일정 기간 동안의 컴퓨터 단층 촬영(computed tomography scan, CT scan)에 대한 방사선학 보고서를 입력 데이터로 사용하여 상기 입력 데이터로부터 췌장암 생존 가능성인 췌장암 생존율을 예측하도록 훈련되는, 훈련 장치.

청구항 12

제11항에서,

상기 췌장암 생존 예측 모델은,

상기 영상 진단 보고서를 자연어 처리하는 상기 언어 모델, 그리고 자연어 처리 결과를 생존 분석하여 상기 췌장암 생존율을 예측하는 생존 분석 모델을 포함하는, 훈련 장치.

청구항 13

제12항에서,

비교 대상인 적어도 두개의 언어 모델 별로 상기 췌장암 생존율의 예측 성능을 산출하여 비교 분석하고, 상기 적어도 두개의 언어 모델 별로 산출된 예측 성능 지표의 비교 분석 정보를 출력하는 모델 평가부

를 더 포함하는, 훈련 장치.

청구항 14

제12항에서,

상기 영상 진단 보고서의 개수를 다양하게 차등 적용한 복수개의 검증 데이터를 이용하여, 상기 췌장암 생존율의 예측 성능을 산출하여 비교 분석하고, 상기 영상 진단 보고서의 개수 별로 산출된 예측 성능 지표의 비교 분석 정보를 출력하는 모델 평가부

를 더 포함하는, 훈련 장치.

청구항 15

제12항에서,

적어도 하나의 특정 단어의 포함 유무에 따라 영상 진단 보고서를 구분하여 검증 데이터를 구성하고, 상기 검증 데이터를 이용하여 상기 특정 단어의 포함 유무에 따른 상기 췌장암 생존율의 예측 성능을 산출하여 비교 분석하고, 상기 영상 진단 보고서의 개수 별로 산출된 예측 성능 지표의 비교 분석 정보를 출력하는 모델 평가부

를 더 포함하는, 훈련 장치.

청구항 16

적어도 하나의 프로세서에 의해 동작하는 훈련 장치의 훈련 방법으로서,

췌장암 환자의 영상 진단 보고서를 학습 데이터로 생성하는 단계, 그리고

상기 학습 데이터를 이용하여 입력 데이터로부터 췌장암 환자의 생존율을 예측하도록 췌장암 생존 예측 모델을 훈련시키는 단계

를 포함하는, 훈련 장치.

청구항 17

제16항에서,

상기 훈련 시키는 단계는,

췌장암 환자에 대한 일정 기간 동안의 컴퓨터 단층 촬영(computed tomography scan, CT scan)에 대한 방사선학 보고서를 학습 데이터로 사용하는, 훈련 장치.

청구항 18

제16항에서,

상기 훈련 시키는 단계는,

상기 영상 진단 보고서를 자연어 처리하는 상기 언어 모델, 그리고 자연어 처리 결과를 생존 분석하여 상기 췌장암 생존율을 예측하는 생존 분석 모델을 훈련시키는, 훈련 장치.

청구항 19

제18항에서,

상기 훈련 시키는 단계 이후,

적어도 하나의 평가 조건을 적용하여, 상기 언어 모델에 대한 췌장암 생존율의 예측 성능 지표를 분석하는 모델 성능 평가를 수행하는 단계를 더 포함하고,

상기 예측 성능 지표는,

예측한 췌장암 생존율과 췌장암 환자의 실제 생존률 간의 일치 지수(concordance index)인 생존 함수(c-index), 예측한 췌장암 생존율과 췌장암 환자의 전체 생존률(overall survival) 간의 연관성 평가 지수인 스피어만 상관 계수, 그리고 췌장암 환자가 1년 이상 생존했는지 예측한 평가 지수인 AUROC(area under the receiver operating characteristic curve) 중 적어도 하나를 포함하는, 훈련 장치.

청구항 20

제19항에서,

상기 모델 성능 평가는,

비교 대상인 적어도 두개의 언어 모델 별로 상기 췌장암 생존율의 예측 성능 지표를 산출하고, 산출한 예측 성능 지표를 비교 분석한 정보를 출력하는, 훈련 장치.

청구항 21

제19항에서,

상기 모델 성능 평가는,

상기 영상 진단 보고서의 개수를 다양하게 차등 적용한 복수개의 검증 데이터를 이용하여, 상기 췌장암 생존율의 예측 성능 지표를 산출하고, 산출한 예측 성능 지표를 비교 분석한 정보를 출력하는, 훈련 장치.

청구항 22

제19항에서,

상기 모델 성능 평가는,

적어도 하나의 특정 단어의 포함 유무에 따라 영상 진단 보고서를 구분하여 검증 데이터를 구성하고, 상기 검증 데이터를 이용하여 상기 특정 단어의 포함 유무에 따른 상기 채장암 생존율을 산출하여 비교 분석하고, 비교 분석 정보를 출력하는, 훈련 장치.

발명의 설명

기술 분야

[0001] 본 발명은 영상 진단 보고서를 이용한 채장암 환자의 생존율 예측 방법, 예측 장치, 훈련 방법 및 훈련 장치에 관한 것이다.

배경 기술

[0002] 채장암은 전 세계적으로 암 관련 사망의 주요 원인 중 하나이며, 그 암울한 예후는 지난 수십 년 동안 거의 개선되지 않았다. 최근 인공지능(AI) 알고리즘의 발전에 따라, 사용 가능한 임상 데이터에서 더 많은 정보를 추출하여 채장암 결과에 대한 이해를 향상시키기 위한 노력이 진행 중이다.

[0003] 방사선학에서의 자연어 처리(NLP) 연구는 대부분 특징의 자동 라벨링을 용이하게 함으로써 "기본 진실"을 확립하는 데 초점을 맞추고 있어, 인공지능 연구에서의 하류 분석을 가속화하는데 도움을 주고 있다.

[0004] 그러나, 자유 형식의 방사선학 보고서에 NLP를 적용하는데 초점을 맞춘 연구는 잠재적으로 추출할 수 있는 정보의 풍부함에도 불구하고 아직은 부족한 실정이다.

발명의 내용

해결하려는 과제

[0005] 본 개시는, 채장암 환자의 컴퓨터 단층 촬영에 대한 방사선학 보고서인 영상 진단 보고서를 기초로, 자연어 처리를 통해 채장암 생존을 예측하는 방법, 훈련 방법, 예측 장치 및 훈련 장치를 제공하고자 한다.

과제의 해결 수단

[0006] 한 특징에 따르면, 적어도 하나의 프로세서에 의해 동작하는 예측 장치의 예측 방법으로서, 채장암 환자에 대한 의사의 임상적 조건이 기재된 영상 진단 보고서를 입력 받는 단계, 그리고 입력 데이터로부터 채장암 생존 가능성인 채장암 생존율을 예측하도록 훈련된 채장암 생존율 예측 모델을 이용하여, 상기 입력 받은 영상 진단 보고서로부터 상기 채장암 환자의 생존율을 예측하는 단계를 포함한다.

[0007] 상기 예측하는 단계는, 상기 영상 진단 보고서를 자연어 처리 모델에 입력하여 상기 영상 진단 보고서에 대한 복수개의 단어 특징 벡터를 추출하고, 추출한 복수개의 단어 특징 벡터들을 통합한 단일 특징 벡터를 보고서 특징 벡터로 추출하며, 추출한 보고서 특징 벡터를 생존 분석 모델에 입력하여 상기 채장암 생존율을 획득할 수 있다.

[0008] 상기 자연어 처리 모델은, ClinicalBERT(Bidirectional Encoder Representations from Transformers) 모델일 수 있다.

[0009] 상기 예측하는 단계는, 상기 자연어 처리 모델을 통해 연속된 복수개의 영상 진단 보고서로부터 추출한 보고서 특징 벡터를 순환 신경망(Recurrent Neural Network, RNN)을 통해 연속적으로 분석하여 환자 임베딩 벡터를 추출하고, 추출한 환자 임베딩 벡터를 상기 생존 분석 모델에 입력하여 상기 채장암 생존율을 획득할 수 있다.

[0010] 상기 생존 분석 모델은, 콕스 비례 위험 모델(Cox proportional-hazards)을 포함하고, 상기 채장암 생존율은, 콕스 비례 위험도(Cox Proportional-Hazard Rate)일 수 있다.

[0011] 다른 특징에 따르면, 적어도 하나의 프로세서에 의해 동작하는 예측 장치로서, 입력 데이터로부터 채장암 생존 가능성인 채장암 생존율을 예측하도록 훈련된 인공지능 모델로서, 입력 받은 채장암 환자에 대한 의사의 임상적

소견이 기재된 영상 진단 보고서로부터 상기 췌장암 환자의 생존율을 예측하는 췌장암 생존율 예측 모델을 포함한다.

- [0012] 상기 췌장암 생존율 예측 모델은, 상기 영상 진단 보고서에 대한 복수개의 단어 특징 벡터를 통합한 단일 특징 벡터를 보고서 특징 벡터로 추출하는 자연어 처리 모델, 그리고 상기 보고서 특징 벡터를 입력 받아 상기 췌장암 생존율을 출력하는 생존 분석 모델을 포함할 수 있다.
- [0013] 상기 영상 진단 보고서는, 췌장암 환자에 대한 일정 기간 동안의 컴퓨터 단층 촬영(computed tomography scan, CT scan)에 대한 방사선학 보고서일 수 있다.
- [0014] 상기 췌장암 생존율 예측 모델은, 상기 자연어 처리 모델에 의해 추출된 연속된 복수개의 영상 진단 보고서로부터 각각의 보고서 특징 벡터를 순차적으로 입력 받고, 입력 받은 보고서 특징 벡터에 대한 연속 분석을 통해 환자 임베딩 벡터를 추출하여 상기 생존 분석 모델에 입력하는 순환 신경망(Recurrent Neural Network, RNN)을 더 포함할 수 있다.
- [0015] 상기 순환 신경망은, GRU(gated rectified unit)를 사용하여 보고서 특징 벡터를 연속적으로 분석해서 환자 임베딩 벡터를 출력할 수 있다.
- [0016] 또 다른 특징에 따르면, 적어도 하나의 프로세서에 의해 동작하는 훈련 장치로서, 췌장암 환자의 영상 진단 보고서를 학습 데이터로 사용하여, 적어도 하나의 태스크를 학습한 췌장암 생존 예측 모델을 생성하는 모델 학습부를 포함하고, 상기 췌장암 생존 예측 모델은, 췌장암 환자에 대한 일정 기간 동안의 컴퓨터 단층 촬영(computed tomography scan, CT scan)에 대한 방사선학 보고서를 입력 데이터로 사용하여 상기 입력 데이터로부터 췌장암 생존 가능성인 췌장암 생존율을 예측하도록 훈련될 수 있다.
- [0017] 상기 췌장암 생존 예측 모델은, 상기 영상 진단 보고서를 자연어 처리하는 상기 언어 모델, 그리고 자연어 처리 결과를 생존 분석하여 상기 췌장암 생존율을 예측하는 생존 분석 모델을 포함할 수 있다.
- [0018] 상기 훈련 장치는, 비교 대상인 적어도 두개의 언어 모델 별로 상기 췌장암 생존율의 예측 성능을 산출하여 비교 분석하고, 상기 적어도 두개의 언어 모델 별로 산출된 예측 성능 지표의 비교 분석 정보를 출력하는 모델 평가부를 더 포함할 수 있다.
- [0019] 상기 훈련 장치는, 상기 영상 진단 보고서의 개수를 다양하게 차등 적용한 복수개의 검증 데이터를 이용하여, 상기 췌장암 생존율의 예측 성능을 산출하여 비교 분석하고, 상기 영상 진단 보고서의 개수 별로 산출된 예측 성능 지표의 비교 분석 정보를 출력하는 모델 평가부를 더 포함할 수 있다.
- [0020] 상기 훈련 장치는, 적어도 하나의 특정 단어의 포함 유무에 따라 영상 진단 보고서를 구분하여 검증 데이터를 구성하고, 상기 검증 데이터를 이용하여 상기 특정 단어의 포함 유무에 따른 상기 췌장암 생존율의 예측 성능을 산출하여 비교 분석하고, 상기 영상 진단 보고서의 개수 별로 산출된 예측 성능 지표의 비교 분석 정보를 출력하는 모델 평가부를 더 포함할 수 있다.
- [0021] 또 다른 특징에 따르면, 적어도 하나의 프로세서에 의해 동작하는 훈련 장치의 훈련 방법으로서, 췌장암 환자의 영상 진단 보고서를 학습 데이터로 생성하는 단계, 그리고 상기 학습 데이터를 이용하여 입력 데이터로부터 췌장암 환자의 생존율을 예측하도록 췌장암 생존 예측 모델을 훈련시키는 단계를 포함한다.
- [0022] 상기 훈련 시키는 단계는, 췌장암 환자에 대한 일정 기간 동안의 컴퓨터 단층 촬영(computed tomography scan, CT scan)에 대한 방사선학 보고서를 학습 데이터로 사용할 수 있다.
- [0023] 상기 훈련 시키는 단계는, 상기 영상 진단 보고서를 자연어 처리하는 상기 언어 모델, 그리고 자연어 처리 결과를 생존 분석하여 상기 췌장암 생존율을 예측하는 생존 분석 모델을 훈련시킬 수 있다.
- [0024] 상기 훈련 시키는 단계 이후, 적어도 하나의 평가 조건을 적용하여, 상기 언어 모델에 대한 췌장암 생존율의 예측 성능 지표를 분석하는 모델 성능 평가를 수행하는 단계를 더 포함하고, 상기 예측 성능 지표는, 예측한 췌장암 생존율과 췌장암 환자의 실제 생존률 간의 일치 지수(concordance index)인 생존 함수(c-index), 예측한 췌장암 생존율과 췌장암 환자의 전체 생존률(overall survival) 간의 연관성 평가 지수인 스피어만 상관 계수, 그리고 췌장암 환자가 1년 이상 생존했는지 예측한 평가 지수인 AUROC(area under the receiver operating characteristic curve) 중 적어도 하나를 포함할 수 있다.
- [0025] 상기 모델 성능 평가는, 비교 대상인 적어도 두개의 언어 모델 별로 상기 췌장암 생존율의 예측 성능 지표를 산출하고, 산출한 예측 성능 지표를 비교 분석한 정보를 출력할 수 있다.

[0026] 상기 모델 성능 평가는, 상기 영상 진단 보고서의 개수를 다양하게 차등 적용한 복수개의 검증 데이터를 이용하여, 상기 체장암 생존율의 예측 성능 지표를 산출하고, 산출한 예측 성능 지표를 비교 분석한 정보를 출력할 수 있다.

[0027] 상기 모델 성능 평가는, 적어도 하나의 특정 단어의 포함 유무에 따라 영상 진단 보고서를 구분하여 검증 데이터를 구성하고, 상기 검증 데이터를 이용하여 상기 특정 단어의 포함 유무에 따른 상기 체장암 생존율을 산출하여 비교 분석하고, 비교 분석 정보를 출력할 수 있다.

발명의 효과

[0028] 본 개시에 따르면, 체장암 환자의 컴퓨터 단층 촬영에 대한 방사선학 보고서인 영상 진단 보고서를 기초로 해당 환자의 생존 여부를 예측할 수 있다.

도면의 간단한 설명

[0029] 도 1은 실시예에 따른 예측 장치를 설명하는 도면이다.
 도 2는 실시예에 따른 예측 장치의 연동 환경을 설명하는 도면이다.
 도 3은 실시예에 따른 예측 방법을 설명하는 도면이다.
 도 4는 실시예에 따른 훈련 장치의 구성도이다.
 도 5는 실시예에 따른 훈련 방법을 설명하는 도면이다.
 도 6은 한 실시예에 따른 모델 평가 방법을 설명하는 도면이다.
 도 7은 다른 실시예에 따른 모델 평가 방법을 설명하는 도면이다.
 도 8은 또 다른 실시예에 따른 모델 평가 방법을 설명하는 도면이다.
 도 9는 실시예에 따른 컴퓨팅 장치의 구성도이다.
 도 10은 실험예에 따른 모델의 예측 성능 지표로서 c-index를 비교 분석한 결과를 나타낸 표이다.
 도 11은 실험예에 따른 회귀분석시 EMR 분석수에 비례해 정확도가 상승하는 결과를 나타낸 그래프이다.
 도 12는 실험예에 따른 모델의 예측 성능 지표로서 AUROC를 비교 분석한 결과를 나타낸 그래프이다.
 도 13은 실험예에 따른 모델의 예측 성능 지표로서 r_s 를 비교 분석한 결과를 나타낸 그래프이다.
 도 14의 (a)는 생존 진단 지수와 환자 그룹 간의 관계를 나타낸 그래프이고, 도 14의 (b)는 환자 그룹 별로 년도 별 생존 확률을 나타낸 그래프이다.
 도 15는 실험예에 따른 모델의 AUROC를 분석한 결과를 나타낸 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0030] 아래에서는 첨부한 도면을 참고로 하여 본 개시의 실시예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 개시는 여러가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 개시를 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0031] 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.

[0032] 또한, 명세서에 기재된 "...부", "...기", "...모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0033] 본 발명에서 설명하는 장치들은 적어도 하나의 프로세서, 메모리 장치, 통신 장치 등을 포함하는 하드웨어로 구성되고, 지정된 장소에 하드웨어와 결합되어 실행되는 프로그램이 저장된다. 하드웨어는 본 발명의 방법을 실행할 수 있는 구성과 성능을 가진다. 프로그램은 도면들을 참고로 설명한 본 발명의 동작 방법을 구현한 명령어

(instructions)를 포함하고, 프로세서와 메모리 장치 등의 하드웨어와 결합하여 본 발명을 실행한다.

- [0034] 본 명세서에서 "전송 또는 제공"은 직접적인 전송 또는 제공하는 것 뿐만 아니라 다른 장치를 통해 또는 우회 경로를 이용하여 간접적으로 전송 또는 제공도 포함할 수 있다.
- [0035] 본 명세서에서 단수로 기재된 표현은 "하나" 또는 "단일" 등의 명시적인 표현을 사용하지 않은 이상, 단수 또는 복수로 해석될 수 있다.
- [0036] 본 명세서에서 도면에 관계없이 동일한 도면번호는 동일한 구성요소를 지칭하며, "및/또는"은 언급된 구성 요소들의 각각 및 하나 이상의 모든 조합을 포함한다.
- [0037] 본 명세서에서, 제1, 제2 등과 같이 서수를 포함하는 용어들은 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되지는 않는다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 개시의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다.
- [0038] 본 명세서에서 도면을 참고하여 설명한 흐름도에서, 동작 순서는 변경될 수 있고, 여러 동작들이 병합되거나, 어느 동작이 분할될 수 있고, 특정 동작은 수행되지 않을 수 있다.
- [0039] 본 개시의 장치는 적어도 하나의 프로세서가 명령어들(instructions)을 실행함으로써, 본 개시의 동작을 수행할 수 있도록 구성 및 연결된 컴퓨팅 장치이다. 컴퓨터 프로그램은 프로세서가 본 개시의 동작을 실행하도록 기술된 명령어들(instructions)을 포함하고, 비일시적-컴퓨터 판독가능 저장매체(non-transitory computer readable storage medium)에 저장될 수 있다. 컴퓨터 프로그램은 네트워크를 통해 다운로드되거나, 제품 형태로 판매될 수 있다.
- [0040] 본 개시의 인공지능 모델(Artificial Intelligence model, AI model)은 적어도 하나의 태스크(task)를 학습하는 기계학습모델로서, 프로세서에 의해 실행되는 컴퓨터 프로그램으로 구현될 수 있다. 인공지능 모델이 학습하는 태스크란, 기계 학습을 통해 해결하고자 하는 과제 또는 기계 학습을 통해 수행하고자 하는 작업을 지칭할 수 있다. 인공지능 모델은 컴퓨팅 장치에서 실행되는 컴퓨터 프로그램으로 구현될 수 있고, 네트워크를 통해 다운로드되거나, 제품 형태로 판매될 수 있다. 또는 인공지능 모델은 네트워크를 통해 다양한 장치들과 연동할 수 있다.
- [0042] 도 1은 실시예에 따른 예측 장치를 설명하는 도면이고, 도 2는 실시예에 따른 예측 장치의 연동 환경을 설명하는 도면이다.
- [0043] 도 1을 참조하면, 예측 장치(Prediction device)(100)는 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치이다.
- [0044] 예측 장치(100)는 췌장암 생존율 예측 모델(10)을 포함할 수 있다.
- [0045] 췌장암 생존율 예측 모델(10)은 입력 데이터로부터 췌장암 생존율 예측을 제공하도록 학습된 인공지능 모델이다.
- [0046] 입력 데이터는 영상 진단 보고서로서, 영상 진단 보고서는 췌장암 환자의 컴퓨터 단층 촬영(computed tomography scan, CT scan)에 대한 방사선학 보고서이다. 방사선학 보고서는 CT 스캔 이미지와는 독립적인 보고서로서, 방사선과 의사들이 임상 해석과 판단에 기초하여 CT 스캔 이미지에 대한 소견을 기재한 텍스트 문서이다. 방사선학 보고서는 주로 의사들 간의 의사 소통에 사용되는 정형화된 용어와 설명으로 구성되어 있지만, 자유롭게 작성된 형태이다.
- [0047] 췌장암 생존율 예측은 췌장암 생존 진단 지수를 출력하며, 췌장암 생존 진단 지수는 콕스 비례 위험도(Cox Proportional-Hazard Rate)일 수 있다.
- [0048] 췌장암 생존율 예측 모델(10)은 복수개의 모델로 구성된다. 췌장암 생존율 예측 모델(10)은 언어 모델(11), 심층 신경망(12) 및 생존 분석 모델(13)을 포함한다.
- [0049] 췌장암 환자들은 치료 반응과 질병 진행 상황을 평가하기 위해 치료 과정 동안 일련의 CT 스캔을 받게 되며, 이러한 과정을 통해 획득한 복수개의 연속된 영상 진단 보고서가 입력 데이터로 활용될 수 있다. 따라서, 췌장암 생존율 예측 모델(10)은 순차적으로 영상 진단 보고서를 언어 모델(11)에 입력하여 추출한 보고서 특징 벡터를

심층 신경망(12)에 입력하여 연속된 분석을 통해 환자 임베딩 벡터를 추출한다.

- [0050] 언어 모델(11)은 영상 진단 보고서를 입력 받아 입력 받은 영상 진단 보고서에 대한 보고서 특징 벡터를 출력한다. 언어 모델(11)은 입력 받은 영상 진단 보고서를 언어 모델(11) 내의 각 단어에 해당하는 임베딩 벡터, 즉, 단어 임베딩 벡터들로 변환하고, 단어 임베딩 벡터들을 통합한 단일 특징 벡터를 보고서 특징 벡터로 출력한다.
- [0051] 이때, 언어 모델(11)은 복수의 연속된 영상 진단 보고서를 입력 받아 연속된 보고서 특징 벡터를 출력할 수 있다.
- [0052] 언어 모델(11)은 자연어 처리 모델로서, BERT(Bidirectional Encoder Representations from Transformers) 모델, ClinicalBERT 모델이 사용될 수 있다. 후술할 모델 성능 평가에 따르면 ClinicalBERT 모델의 예측 성능 지표가 상대적으로 높으므로, 언어 모델(11)은 ClinicalBERT 모델이 사용될 수 있다.
- [0053] 심층 신경망(12)은 순환 신경망(Recurrent Neural Network, RNN)으로서, 연속적인 영상 진단 보고서로부터 출력되는 보고서 특징 벡터를 연속적으로 분석해서 환자 임베딩 벡터를 출력한다. 심층 신경망(12)은 모든 반복적인 RNN 중에서 GRU(gated rectified unit)를 사용하여 보고서 특징 벡터를 연속적으로 분석해서 환자 임베딩 벡터를 출력한다.
- [0054] 생존 분석 모델(13)은 생존 분석을 통해 환자 임베딩 벡터로부터 환자의 생존 위험도인 생존 진단 지수를 출력한다. 생존 분석 모델(13)은 콕스 비례 위험 모델(Cox proportional-hazards)이 사용될 수 있으며 생존 진단 지수는 콕스 비례 위험도(Cox Proportional-Hazard Rate)이다.
- [0055] 도 2를 참조하면, 예측 장치(100)는 독립된 장치이거나 또는 다른 장치들과의 연동형 장치로 구현될 수 있다.
- [0056] 예측 장치(100)는 EMR(Electronic Medical Records) 시스템(200) 또는 디바이스(300)로부터 환자의 영상 진단 보고서를 입력 받을 수 있다.
- [0057] 디바이스(300)는 예측 장치(100) 및 EMR 시스템(200)과 연동하여 필요한 정보를 화면에 표시하는 사용자 인터페이스를 제공할 수 있다.
- [0058] 예측 장치(100)는 디바이스(300)로부터 웨장암 환자의 생존율 예측을 요청 받고, 디바이스(300)에게 요청된 예측에 응답할 수 있다. 디바이스(300)는 예측 장치(100)로의 접근 권한이 부여된 의료기관 단말, 예컨대, 의사 단말일 수 있다.
- [0059] 예측 장치(100)는 서버 장치일 수 있고, 디바이스(300)는 의료기관에 설치된 클라이언트 단말일 수 있으며, 이들(100, 300)은 네트워크를 통해 연동할 수 있다.
- [0060] 예측 장치(100)는 특정 의료 기관 내 네트워크에 연결된 로컬 서버일 수 있다.
- [0061] 예측 장치(100)는 클라우드 서버일 수 있고, 접근 권한을 가진 다수의 의료 기관 단말(의사 단말)과 연동할 수 있다.
- [0063] 도 3은 한 실시예에 따른 예측 방법을 설명하는 도면으로서, 도 1, 2의 구성과 연계하여 동일한 구성에 대해 동일한 도면 부호를 사용하여 설명한다.
- [0064] 도 3을 참고하면, 예측 장치(100)는 웨장암 환자의 영상 진단 보고서를 입력 받는다(S101).
- [0065] 예측 장치(100)는 입력 받은 영상 진단 보고서의 누적 개수가 임계 개수를 충족하는지 판단한다(S102).
- [0066] 예측 장치(100)는 입력 받은 영상 진단 보고서의 누적 개수가 임계 개수를 충족하지 않으면, 누적 개수만큼의 영상 진단 보고서를 분석 대상으로 선택한다(S103).
- [0067] 예측 장치(100)는 입력 받은 영상 진단 보고서의 누적 개수가 임계 개수를 충족하면, 임계 개수만큼의 영상 진단 보고서를 분석 대상으로 선택한다(S104).
- [0068] 예측 장치(100)는 언어 모델(11), 심층 신경망(12) 및 생존 분석 모델(13)을 이용하여, S103 또는 S104에서 선택한 개수의 영상 진단 보고서들을 토대로, 생존 진단 지수를 획득하는 절차를 진행한다.
- [0069] 먼저, 예측 장치(100)는 개별 영상 진단 보고서를 언어 모델에 입력하여 언어 모델로부터 복수의 단어 임베딩 벡터를 생성하고, 생성한 복수의 단어 임베딩 벡터 중에서 상대적으로 중요한 단어 임베딩 벡터에 가중치를 부

여하고 집계하여 단일 특징 벡터, 즉, 보고서 특징 벡터를 획득한다(S105).

- [0070] 예측 장치(100)는 S105의 보고서 특징 벡터를 심층 신경망(12)에 입력하여 환자 임베딩 벡터를 획득한다(S106).
- [0071] 도 1에서 설명한 바와 같이, S105와 S106의 과정은 임계 개수 또는 누적 개수만큼의 영상 진단 보고서에 대해 각각 수행되고, 심층 신경망(12)의 출력은 다음 영상 진단 보고서의 심층 신경망(12)의 입력으로 함께 사용된다.
- [0072] 예측 장치(100)는 임계 개수 또는 누적 개수만큼의 영상 진단 보고서들에 대한 S105와 S106의 연속된 절차들을 수행하고, 심층 신경망(12)의 연속된 분석, 즉, 순환 신경망 분석을 통해 최종 환자 임베딩 벡터를 획득한다(S107).
- [0073] 예측 장치(100)는 최종 환자 임베딩 벡터를 입력 데이터로 사용하여 생존 분석 모델(13)을 통해 생존 진단 지수를 획득한다(S108). 생존 진단 지수가 췌장암 환자의 생존 예측률로 출력된다.
- [0075] 도 4는 실시예에 따른 훈련 장치의 구성도이다.
- [0076] 도 4를 참조하면, 훈련 장치(Training Apparatus)(400)는 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치이다.
- [0077] 훈련 장치(400)는 모델 학습부(410) 및 모델 평가부(420)을 포함할 수 있다.
- [0078] 모델 학습부(410)는 췌장암 환자의 영상 진단 보고서가 저장된 훈련 데이터베이스(500)와 연동하고, 훈련 데이터베이스(500)의 영상 진단 보고서를 이용하여 적어도 하나의 태스크를 학습한 췌장암 생존율 예측 모델(10)을 출력한다.
- [0079] 모델 학습부(410)는 췌장암 생존율 예측 모델(10)에서 RNN(12) 및 생존 분석 모델(13)이 입력 데이터로부터 췌장암 생존율을 출력하도록 훈련 시킨다.
- [0080] 모델 평가부(420)는 모델 학습부(410)에 의해 학습된 모델의 예측 성능을 평가한다. 모델 평가부(420)는 다양한 예측 성능 지표를 출력할 수 있는데, 실시예에 따르면, c-index, r_s , AUROC를 출력할 수 있다.
- [0081] c-index는 췌장암 생존율 예측 모델(10)이 예측한 췌장암 생존율과 췌장암 환자의 실제 생존률 간의 일치 지수(concordance index)인 생존 함수를 나타낸다.
- [0082] 스피어만 상관 계수(Spearman's correlation coefficient)(r_s)는 췌장암 생존율 예측 모델(10)이 예측한 췌장암 생존율과 췌장암 환자의 전체 생존률(overall survival) 간의 연관성 평가 지수이다.
- [0083] AUROC(area under the receiver operating characteristic curve)는 환자가 1년 이상 생존했는지 예측한 평가 지수이다.
- [0084] 모델 평가부(420)는 췌장암 생존율 예측 모델(10)의 성능을 평가한다. 모델 평가부(420)는 모델 평가 결과를 출력할 수 있다.
- [0085] 모델 평가부(420)는 비교 대상인 적어도 두개의 언어 모델 별로 췌장암 생존율의 예측 성능을 산출하여 비교 분석하고, 적어도 두개의 언어 모델 별로 산출된 예측 성능 지표의 비교 분석 정보를 출력할 수 있다.
- [0086] 모델 평가부(420)는 영상 진단 보고서의 개수를 다양하게 차등 적용한 복수개의 검증 데이터를 이용하여, 췌장암 생존율의 예측 성능을 산출하여 비교 분석하고, 영상 진단 보고서의 개수 별로 산출된 예측 성능 지표의 비교 분석 정보를 출력할 수 있다.
- [0087] 모델 평가부(420)는 적어도 하나의 특정 단어의 포함 유무에 따라 영상 진단 보고서를 구분하여 검증 데이터를 구성하고, 검증 데이터를 이용하여 특정 단어의 포함 유무에 따른 췌장암 생존율의 예측 성능을 산출하여 비교 분석하고, 영상 진단 보고서의 개수 별로 산출된 예측 성능 지표의 비교 분석 정보를 출력할 수 있다.
- [0089] 도 5는 실시예에 따른 훈련 방법을 설명하는 도면이다.
- [0090] 도 5를 참조하면, 모델 학습부(410)는 췌장암 환자의 연속된 복수개의 영상 진단 보고서에 사망 또는 관찰 종료

시점 까지의 시간을 포함하는 생존 정보를 레이블로 설정한 학습 데이터를 구성한다(S201).

- [0091] 모델 학습부(410)는 S201에서 생성한 학습 데이터를 이용하여, 입력 데이터로부터 생존 진단 지수를 출력하도록 신경망(도 1의 12) 및 생존 분석 모델(도 1의 13)을 훈련한다(S202).
- [0093] 도 6은 한 실시예에 따른 모델 평가 방법을 설명하는 도면이다.
- [0094] 도 6을 참조하면, 모델 평가부(420)는 예측 성능 비교를 위한 복수개의 언어 모델을 선정한다(S301). 실시예에 따르면, 모델 평가부(420)는 BERT 모델과 ClinicalBERT 모델을 평가 대상으로 설정할 수 있다.
- [0095] 모델 평가부(420)는 모델 학습부(410)와 연동하여, S301에서 선정한 각 언어 모델을 적용하여 모델 훈련이 진행 되도록 한다(S302).
- [0096] 모델 평가부(420)는 모델 학습부(410)와 연동하여, 검증을 위해 선택된 환자 별 복수의 영상 진단 보고서로 구성된 검증 데이터와 환자의 생존 정보를 학습 데이터로 사용하고, 각각의 언어 모델을 포함하는 체장암 생존율 예측 모델을 이용하여, 검증 데이터로부터 각각의 생존 진단 지수를 획득한다(S303, S304).
- [0097] 모델 평가부(420)는 서로 다른 언어 모델, 즉, BERT 모델과 ClinicalBERT 모델을 적용하여 획득한 각각의 생존 진단 지수와 입력받은 환자의 생존 정보를 이용하여, 적어도 하나의 예측 성능 지표를 산출한다(S305).
- [0098] 모델 평가부(420)는 서로 다른 언어 모델 별로 예측 성능 지표를 비교한 분석 정보를 출력한다(S306).
- [0099] 즉, 모델 평가부(420)는 언어 모델로서 BERT 모델을 사용하였을 때 산출한 생존 진단 지수, 언어 모델로서 ClinicalBERT 모델을 사용하였을 때 산출한 생존 진단 지수를 기초로, 예측 성능 지표, 즉, c-index를 산출할 수 있다.
- [0100] C-index가 높을 수록 예측 성능이 상대적으로 좋을 수 있다. 실험에 따르면, ClinicalBERT 모델은 체장암 환자의 생존 예측에서 원래의 BERT 모델보다 우수함이 증명되었다.
- [0101] 모델 평가부(420)의 언어 모델 평가를 통해 상대적인 성능이 좋다고 평가된 언어 모델, 예컨대, ClinicalBERT 모델이 체장암 생존율 예측 모델(10)에 포함될 수 있다.
- [0103] 도 7은 다른 실시예에 따른 모델 평가 방법을 설명하는 도면이다.
- [0104] 도 7을 참조하면, 모델 평가부(420)는 모델 학습부(410)와 연동하여 영상 진단 보고서의 개수를 차등 설정한 비교 대상인 검증 데이터와 환자의 생존 정보를 학습 데이터로 사용(S401)하여, 체장암 생존율 예측 모델을 이용하여, 비교 대상인 검증 데이터로부터 각각의 생존 진단 지수를 획득한다(S402).
- [0105] 모델 평가부(420)는 각각의 생존 진단 지수와 입력받은 환자의 생존 정보를 이용하여, 적어도 하나의 예측 성능 지표를 산출한다(S403).
- [0106] 모델 평가부(420)는 영상 진단 보고서의 개수 별로 예측 성능 지표를 비교한 분석 정보를 출력한다(S404).
- [0107] 실험에 따르면, 환자당 입력된 CT 보고서의 수가 많을수록 예측 성능이 높아짐이 증명되었다.
- [0109] 도 8은 또 다른 실시예에 따른 모델 평가 방법을 설명하는 도면이다.
- [0110] 도 8을 참조하면, 모델 평가부(420)는 영상 진단 보고서와 관련된 단어들 중에서 환자 생존과 관련된 것으로 예측되는 단어들을 입력 받는다(S501).
- [0111] 모델 평가부(420)는 모델 학습부(410)와 연동하여 검증을 위해 선택된 환자 별 복수의 영상 진단 보고서로 구성된 검증 데이터와 환자의 생존 정보를 학습 데이터로 구성(S502)하도록 하고, 입력된 단어들이 포함된 검증 데이터와 포함되지 않은 검증 데이터로 구분하도록 한다(S503).
- [0112] 모델 평가부(420)는 모델 학습부(410)와 연동하여, S503에서 구분한 검증 데이터를 체장암 생존율 예측 모델에 입력하여 각각의 생존 진단 지수를 획득하고(S504), 각각의 생존 진단 지수와 입력 받은 환자의 생존 정보를 이용하여, 적어도 하나의 예측 성능 지표를 산출한다(S505).

- [0113] 모델 평가부(420)는 입력된 단어들의 포함 유무에 따른 예측 성능 지표를 비교한 분석 정보를 출력한다(S506). 실험에 따르면, 다른 용어와 달리 "liver metastasis/no evidence of"가 영상 진단 보고서에 포함되어 있을 때, 췌장암 환자의 생존율은 상승되었음이 증명되었다.
- [0115] 도 9는 실시예에 따른 컴퓨팅 장치의 구성도이다.
- [0116] 도 9를 참조하면, 도 1 ~ 도 8에서 설명한 예측 장치(100), 훈련 장치(400), 그리고 훈련 장치(400)를 구성하는 모델 학습부(410)와 모델 평가부(420), 디바이스(300)는 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치(600)로 구현될 수 있다.
- [0117] 컴퓨팅 장치(600)는 하나 이상의 프로세서(610), 프로세서(610)에 의하여 수행되는 컴퓨터 프로그램을 로드 하는 메모리(630), 컴퓨터 프로그램 및 각종 데이터를 저장하는 저장 장치(650), 통신 인터페이스(670), 그리고 이들을 연결하는 버스(690)를 포함할 수 있다. 이외에도, 컴퓨팅 장치(600)는 다양한 구성 요소가 더 포함될 수 있다. 프로세서(610)는 컴퓨팅 장치(600)의 동작을 제어하는 장치로서, 컴퓨터 프로그램에 포함된 명령어들을 처리하는 다양한 형태의 프로세서일 수 있고, 예를 들면, CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit), GPU(Graphic Processing Unit) 또는 본 개시의 기술 분야에 잘 알려진 임의의 형태의 프로세서 중 적어도 하나를 포함하여 구성될 수 있다.
- [0118] 메모리(630)는 각종 데이터, 명령 및/또는 정보를 저장한다. 메모리(630)는 본 개시의 동작을 실행하도록 기술된 명령어들이 프로세서(610)에 의해 처리되도록 해당 컴퓨터 프로그램을 저장 장치(650)로부터 로드 할 수 있다. 메모리(630)는 예를 들면, ROM(read only memory), RAM(random access memory) 등 일 수 있다.
- [0119] 저장 장치(650)는 컴퓨터 프로그램, 각종 데이터를 비임시적으로 저장할 수 있다. 저장 장치(650)는 ROM(Read Only Memory), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM), 플래시 메모리 등과 같은 비휘발성 메모리, 하드 디스크, 착탈형 디스크, 또는 본 개시가 속하는 기술 분야에서 잘 알려진 임의의 형태의 컴퓨터로 읽을 수 있는 기록 매체를 포함하여 구성될 수 있다.
- [0120] 통신 인터페이스(670)는 유/무선 통신을 지원하는 유/무선 통신 모듈일 수 있다.
- [0121] 버스(690)는 컴퓨팅 장치(600)의 구성 요소 간 통신 기능을 제공한다.
- [0122] 컴퓨터 프로그램은, 프로세서(610)에 의해 실행되는 명령어들(instructions)을 포함하고, 비일시적-컴퓨터 판독 가능 저장매체(non-transitory computer readable storage medium)에 저장되며, 명령어들은 프로세서(610)가 본 개시의 동작을 실행하도록 만든다. 컴퓨터 프로그램은 네트워크를 통해 다운로드 되거나, 제품 형태로 판매될 수 있다.
- [0123] 컴퓨터 프로그램은, 도 3의 단계들, 도 5의 단계들, 도 6의 단계들, 도 7의 단계들, 도 8의 단계들을 실행하는 명령어들을 포함할 수 있다.
- [0125] 이하에서는 실험예를 통하여 본 발명을 상세하게 설명한다. 이러한 실험예는 단지 본 발명을 예시하기 위한 것이며, 본 발명이 여기에 한정되는 것은 아니다.
- [0126] **실험예**
- [0127] NLP(natural language processing) 모델과 췌장암 환자에 대한 방사선학 보고서를 학습 데이터를 사용하여, 췌장암 환자의 생존율을 예측하는 모델을 개발하였다. NLP 모델은 자유 형식의 방사선학 보고서만을 사용하고, 다른 임상 정보 없이 췌장암 환자의 생존을 예측할 수 있으며, 이러한 예측 정보는 임상 의사결정을 지원할 수 있다.
- [0128] NLP 모델로서, BERT 모델과 ClinicalBERT 모델이 사용되었다.
- [0129] 모델 평가는 컴퓨터 단층 촬영의 방사선학 보고서의 자연어 처리(NLP)만을 기반으로 한 췌장암 생존 예측의 실행 가능성을 확인하기 위한 목적을 가지고 이루어졌다.
- [0130] 딥 러닝 기반의 자연어 처리(NLP) 모델이 3,347명의 췌장암 환자로부터 나온 15,313건의 연속적인 컴퓨터 단층 촬영(CT) 보고서에 적용되었다.

- [0131] 최적의 NLP 모델을 식별하기 위해 일치 지수(c-index)가 계산되었다.
- [0132] 스피어만 상관 계수(r_s)는 모델이 예측한 생존률과 환자의 전체 생존률 사이에 계산되었다.
- [0133] 수신자 운영 특성 곡선 아래의 영역(AUROC)은 1년 생존률을 예측하기 위해 계산되었다.
- [0134] 환자 당 CT 보고서의 다양한 수량으로 훈련된 여러 모델을 비교하여 연속적인 CT 보고서의 높은 수량이 우월함을 입증했다. 일반화를 검증하기 위해, 외국 기관에서 300명의 췌장암 환자로부터의 1,947개의 연속 CT 보고서가 사용되었다. 주요 결과는 예측된 생존률과 췌장암 환자의 실제 생존률의 일치 지수 및 환자 당 연속 CT 보고서 수가 증가함에 따라 계산된 AUROC였다.

[0136] **학습 데이터 세팅**

- [0137] 2004년 1월부터 2021년 6월까지 한국 서울의 연세 암센터에서 진료받은 환자 중에서 3,347명의 췌장암 환자가 선택되었다. 외부 검증을 위해 300명의 환자들이 미국 코네티컷 주의 예일 뉴 헤이븐 병원에서 본 환자들 중에서 선택되었다. 연세 암센터와 뉴 헤이븐 병원 모두 3차 의료기관이다.
- [0138] 3차 의료기관인 연세 암센터의 전자 의료 기록 (EMR) 시스템을 회고적으로 검색하여 2004년 1월부터 2021년 6월까지 췌장암 환자를 연속적으로 수집하고, 유효한 생존 정보가 없는 환자들은 제외되었다.
- [0139] 복부 컴퓨터 단층 촬영(CT) 스캔의 방사선학 보고서는 NLP 모델 개발을 위해 사용되었다. 방사선학 보고서는 25명 이상의 다른 방사선 전문의들에 의해 영어로 작성되었으며, 특정한 구조화된 형태 없이 서술 형식으로 되어 있다.
- [0140] 환자들의 4/5는 훈련 데이터 셋에 할당되었고, 나머지 1/5는 내부 검증 데이터셋을 위해 예약되었다. 훈련 및 내부 검증 세트를 나눌 때, 데이터 세트 간에 환자 당 CT 검사 횟수를 비슷하게 유지하였다.
- [0141] NLP 모델에 의해 예측된 환자 생존과 실제 관찰된 전체 생존(OS) 사이의 일치성을 가지도록 훈련되었으며, 일치성은 초기 CT 검사의 날짜부터 사망 또는 마지막 추적 방문(검열)까지의 날짜로 계산되었다.

[0143] **모델 성능의 내부 검증**

- [0144] 제안된 모델은 다음의 네 단계를 기반으로 환자의 생존 진단 지수를 예측했다.
- [0145] 첫 번째 단계는 초기 방사선학 보고서를 BERT 언어 모델 내의 각 단어에 해당하는 임베딩 벡터로 변환하는 것이며, 이 모델은 입력 텍스트를 단어의 순서로 이해한다.
- [0146] 두 번째 단계는 단어 임베딩 벡터를 단일 특징 벡터로 집계하여 방사선학 보고서를 나타내는 것이고, 이는 보고서에서 더 중요한 단어에 가중치를 주는 주의 기법이 사용되었다.
- [0147] 세 번째 단계는 모든 반복적인 신경망(RNN) 구현 중에서 게이트가 있는 정류 유닛을 사용하여 단일 환자의 여러 방사선학 보고서를 연속적으로 분석하고 환자 임베딩 벡터를 얻는 것이다.
- [0148] 마지막으로, 췌장암 환자의 생존율은 환자 임베딩 벡터를 사용한 콕스 비례 위험 모델로 예측되었다.
- [0149] NLP 모델의 성능을 평가하기 위해, 세 가지 비교 분석이 수행되었다.
- [0150] 먼저, 생존 예측에 대한 성능은 훈련되지 않은 원래의 BERT 모델과 ClinicalBERT 모델 사이에서 비교되었다. 각 모델은 훈련 데이터셋을 사용하여 5-겹 교차 검증으로 훈련되었으며, 모델의 성능은 예약된 내부 검증 데이터셋을 사용하여 비교되었다.
- [0151] 둘째로, 초기 방사선학 보고서만 사용하는 것에 비해 여러 차례의 연속적인 방사선학 보고서를 사용했을 때 모델 성능이 향상되는지 평가했다.
- [0152] 이러한 비교 분석 결과는 도 10 ~ 도 13을 통해 증명되었다.
- [0153] 도 10은 실험예에 따른 모델의 예측 성능 지표로서 c-index를 비교 분석한 결과를 나타낸 표이고, 도 11은 실험예에 따른 회귀분석시 EMR 분석수에 비례해 정확도가 상승하는 결과를 나타낸 그래프이며, 도 12는 실험예에 따른 모델의 예측 성능 지표로서 AUROC를 비교 분석한 결과를 나타낸 그래프이고, 도 13은 실험예에 따른 모델의

예측 성능 지표로서 r_s 를 비교 분석한 결과를 나타낸 그래프이다.

- [0154] NLP 모델의 성능은 일치 지수(C-index)를 사용하여 전체 생존율을 예측하고 수신자 운영 특성 곡선(AUROC) 아래 영역을 사용하여 환자가 1년 이상 생존했는지 예측하는 데 사용되어 평가되었다.
- [0155] 도 10을 참조하면, 단일 CT 보고서를 사용하였을 때, BERT의 c-index는 0.621이고, ClinicalBERT의 c-index는 0.653이다. 따라서, ClinicalBERT는 췌장암 환자의 생존 예측에서 원래의 BERT보다 우수함을 나타낸다.
- [0156] 또한, ClinicalBERT를 대상으로 CT 보고서의 개수를 단일로 하였을 때, 10개개 하였을 때, 15개로 하였을 때 각각의 c-index는 0.653, 0.779, 0.811이다. 따라서, 환자당 입력된 CT 보고서의 수가 많을수록 예측 성능이 높아짐을 알 수 있다.
- [0157] 도 11을 참조하면, CT 보고서의 개수의 증가에 따른 c-index와의 관계를 그래프로 나타냈으며, 이는 다음의 수학적식으로 표현될 수 있다.
- [0158] [수학적식 1]
- [0159] $0.6325 \leq 0.0119x + 0.6514 \leq 0.6800$
- [0160] 즉, 회귀분석시 EMR 분석수에 비례해서 생존 예측 정확도(1이 100%)가 상승함을 나타낸다.
- [0161] 도 12를 참조하면, 단일 CT 보고서를 사용하였을 때, BERT의 AUROC는 0.679이고, 단일 CT 보고서를 사용하였을 때 ClinicalBERT 모델의 AUROC는 0.722, 10개의 CT 보고서를 사용하였을 때 ClinicalBERT 모델의 AUROC는 0.841이고, 15개의 CT 보고서를 사용하였을 때 ClinicalBERT 모델의 AUROC는 0.911이다.
- [0162] 이와 같이, ClinicalBERT는 췌장암 환자의 생존 예측에서 원래의 BERT보다 우수함을 나타낸다. 또한, 환자당 입력된 CT 보고서의 수가 많을수록 예측 성능이 높아짐을 알 수 있다.
- [0163] 도 13을 참조하면, 예측 지수와 관찰된 OS 사이의 연관성을 평가하기 위해 스피어만 상관 계수(r_s)가 사용되었다. 상관계수(r_s)는 -0.79이 도출되었다.
- [0164] 훈련된 ClinicalBERT 모델을 간격 검증을 위해 예약된 환자의 20% (670명의 환자)에 적용하였고, 관찰된 전체 생존률과 예측된 생존 지수 사이에 높은 상관 관계(r_s : -0.79)를 관찰하였다.
- [0165] 상관계수(r_s)의 절대값이 클수록 예측값과 실제값 간의 선형 관계가 성립하며, 성능이 좋음을 나타낸다.
- [0167] 모델이 예측한 생존 진단 지표를 기반으로 환자들을 세 개의 하위 그룹(고위험, 중위험, 저위험 그룹)으로 나누었다. 고위험군은 예측 지수가 상위 33퍼센타일에 속하는 환자들로 정의되었으며, 저위험군은 예측 지수가 하위 33퍼센타일에 속하는 환자들이었다. 이러한 하위 그룹으로 환자들 이 성공적으로 계층화되었는지 평가를 수행하였다.
- [0168] 도 14의 (a)는 생존 진단 지수와 환자 그룹 간의 관계를 나타낸 그래프이고, 도 14의 (b)는 환자 그룹 별로 년도 별 생존 확률을 나타낸 그래프이며, 도 15는 실험예에 따른 모델의 AUROC를 분석한 결과를 나타낸 그래프이다.
- [0169] 도 14의 (a)를 참조하면, 모델이 예측한 생존 진단 지표에 기초하여 고위험, 중간위험, 저위험 그룹으로 간격 검증 세트를 계층화 했을 때, 도 14의 (b)를 참조하면, 명확한 생존 차이가 관찰되었다.
- [0170] 생존률은 저위험군(중앙 OS: 20.2개월; 95% 신뢰구간 [CI]: 19.0, 21.5)에서 중위험군(중앙 OS: 7.3개월; 95% CI: 6.8, 7.6; p -value < 0.001) 보다 좋다.
- [0171] 반면에 고위험군(중앙 OS: 2.6개월; 95% CI: 2.5, 2.9)은 중위험군보다 생존률이 더 나쁘다(p -value < 0.001; 0.001).

[0173] 모델 성능의 외부 검증

- [0174] NLP 모델의 성능을 외부에서 검증하기 위해, 2011년 7월부터 2022년 3월까지 해외 3차 병원(Yale New Haven Hospital, USA)에서 진단된 300명의 췌장암 환자들로부터 얻은 1,947개의 CT 보고서를 사용하였다. EMR은 복부

CT 스캔의 방사선학 보고서와 생존 정보를 포함하고, 이 보고서들은 서술적 스타일로 영어로 작성되었다. 모델의 성능은 전술한 모델 성능의 내부 검증과 같은 방식으로 평가되었다.

[0176] **방사선학 보고서 용어가 환자 생존 예측에 미치는 영향 조사**

[0177] 환자 생존과 관련될 것으로 생각되는 용어들이 선택되었고, 이들이 방사선학 보고서에 포함되어 있는지와 환자 결과 사이의 관계가 조사되었다.

[0178] 방사선학 보고서의 각 용어가 있는 환자 그룹과 없는 환자 그룹을 Kaplan-Meier 플롯을 사용하여 비교하였고, 통계적 유의성은 로그-랭크 검정으로 평가하였다. 게다가, 이러한 용어의 존재가 모델의 환자 생존 예측 능력에 어떤 영향을 미쳤는지 조사하기 위해, 방사선학 보고서에 각 용어가 있는 환자 그룹과 없는 환자 그룹에 대한 모델의 예측 결과를 비교하였다.

[0179] 도 15는 용어와 체장암 생존율 간의 관계를 나타낸 그래프이다.

[0180] 도 15의 A를 참조하면, 용어 '유방암(carcinomatosis)'에 대해 년도에 따른 생존 확률을 나타낸다.

[0181] 도 15의 B를 참조하면, 용어 '간 전이(liver metasta)'에 대해 년도에 따른 생존 확률을 나타낸다.

[0182] 도 15의 C를 참조하면, 용어 '간질환 전이(hepatic metasta)'에 대해 년도에 따른 생존 확률을 나타낸다.

[0183] 도 15의 D를 참조하면, 용어 '원격 전이(distant metasta)'에 대해 년도에 따른 생존 확률을 나타내고, 도 15의 E를 참조하면, 용어 'distant metasta'가 포함된 문장 들의 예시이다.

[0184] 이와 같이, "유방암," "간 전이," "간질환 전이," "원격 전이,"라는 용어가 자유 텍스트 방사선학 보고서에 나타난 환자와 그렇지 않은 환자를 그룹화하여 생존 확률을 산출하여 개별 용어가 생존 확률을 구분할 수 있는지 확인하였다.

[0185] "원격 전이"의 출현은 다른 용어에 비해 더 나은 생존율과 연관됨이 확인되었다. 이 결과는 ClinicalBERT 모델이 전체 맥락을 정확하게 이해할 수 있었음을 보여주며, 특정 구문의 출현에 의존하지 않음을 보여준다.

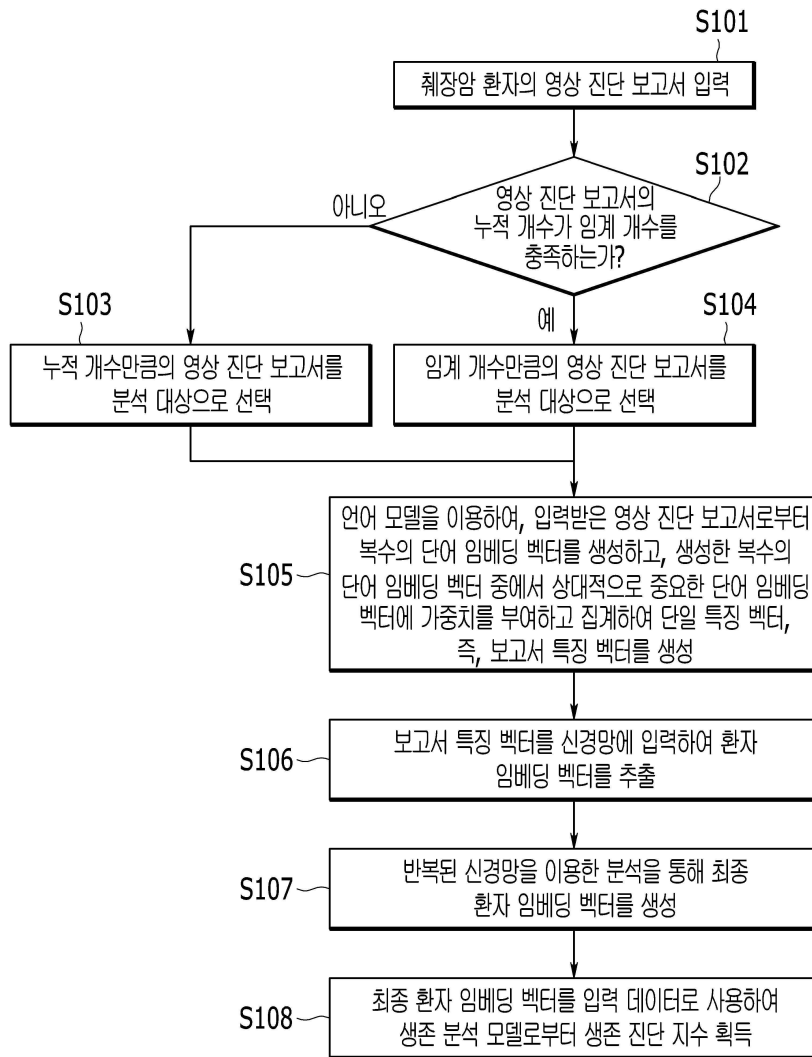
[0186] 도 15의 F를 참조하면, 용어 'no evidence of'에 대해 년도에 따른 생존 확률을 나타낸다. 'no evidence of'의 출현이 생존과 관련이 있는지 추가로 조사하였고, 실제로 이것이 더 나은 생존과 관련이 있음을 발견하였다.

[0187] 이러한 결과에 따르면, ClinicalBERT는 문맥을 정확하게 이해하며 방사선학 보고서의 특정 구문에 의존하지 않고, 모델 성능은 견고하며 특정 용어에 독립적임을 알 수 있다. 즉, 모델이 특정 구문의 출현에만 의존하지 않고, 이러한 구문이 나타나는 맥락을 해석할 수 있음을 나타낸다.

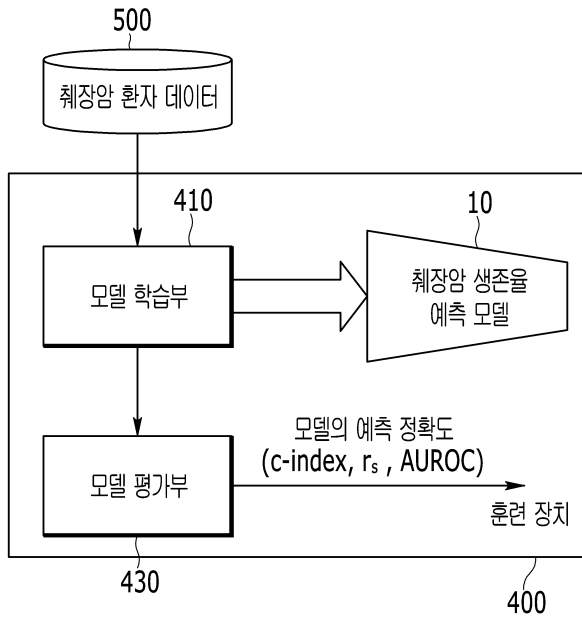
[0189] 이상에서 설명한 본 발명의 실시예는 장치 및 방법을 통해서만 구현이 되는 것은 아니며, 본 발명의 실시예의 구성에 대응하는 기능을 실현하는 프로그램 또는 그 프로그램이 기록된 기록 매체를 통해 구현될 수도 있다.

[0190] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

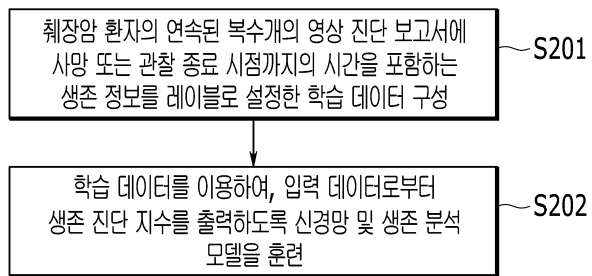
도면3



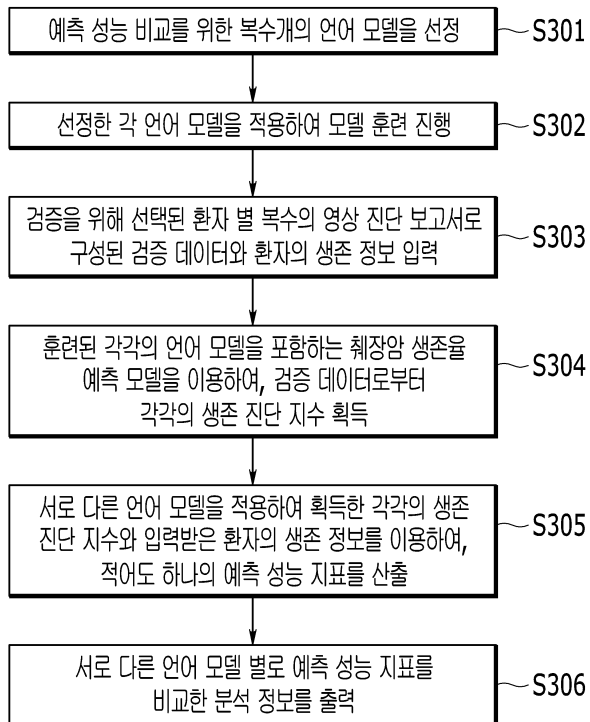
도면4



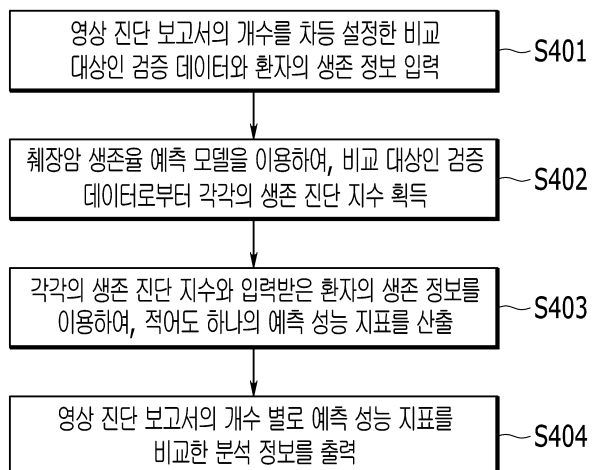
도면5



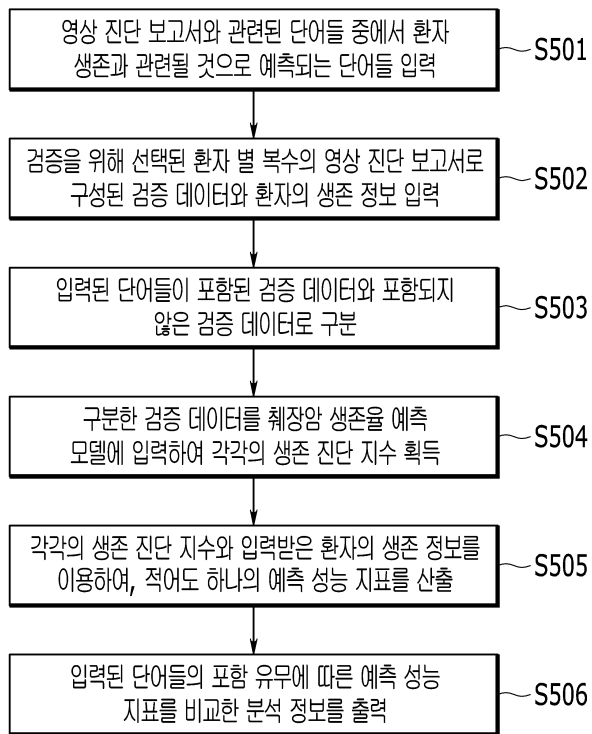
도면6



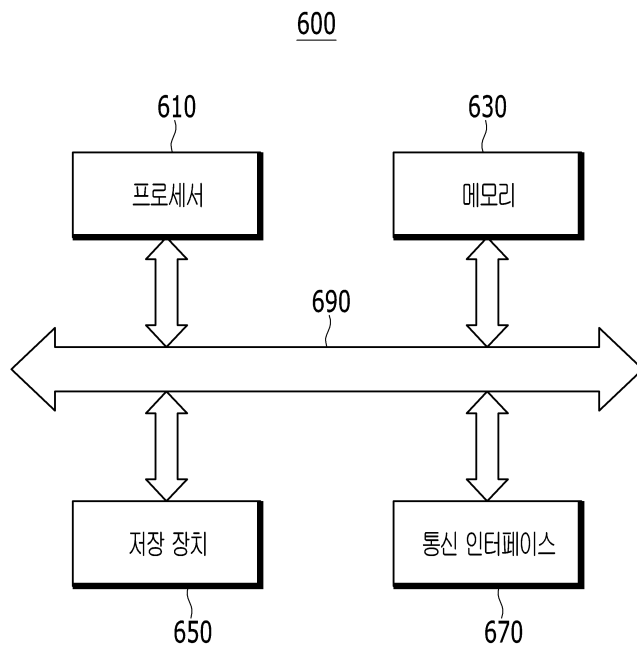
도면7



도면8



도면9

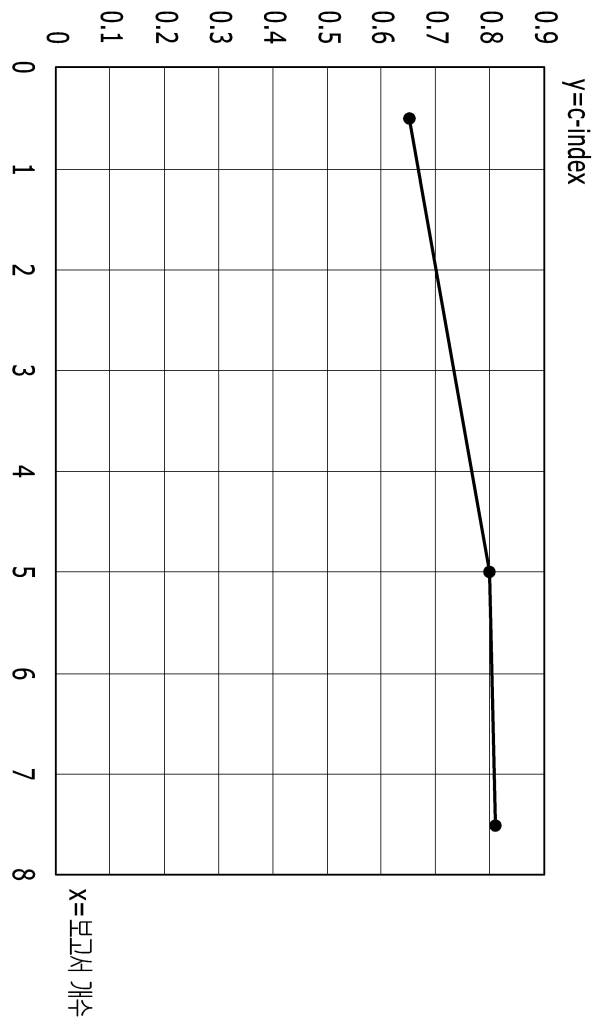


도면10

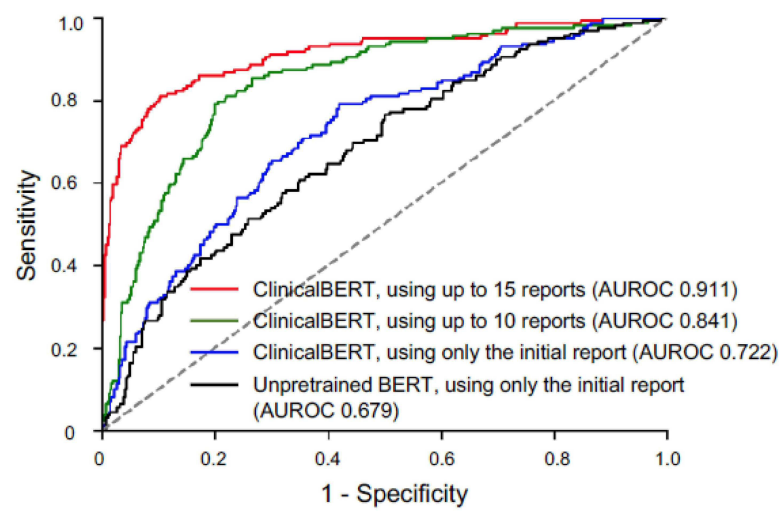
Model	Training					Internal validation set
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	
Using only the first report & unpretrained BERT	0.638	0.579	0.582	0.617	0.616	0.621
Using only the first report & ClinicalBERT	0.655	0.606	0.625	0.644	0.629	0.653
Using up to 10 serial reports & ClinicalBERT	0.802	0.755	0.783	0.786	0.783	0.779
Using up to 15 serial reports & ClinicalBERT (Final model0	0.852	0.771	0.785	0.799	0.791	0.811

BERT, Bidirectional Encoder Representations from Transformers

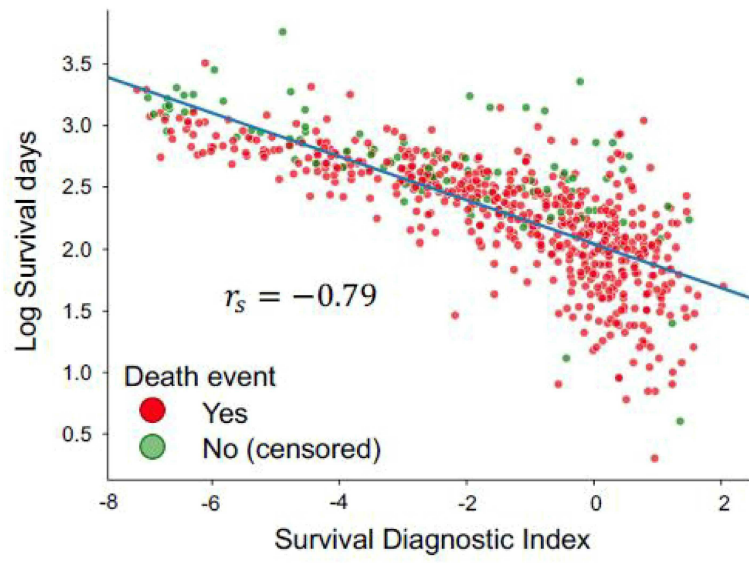
도면11



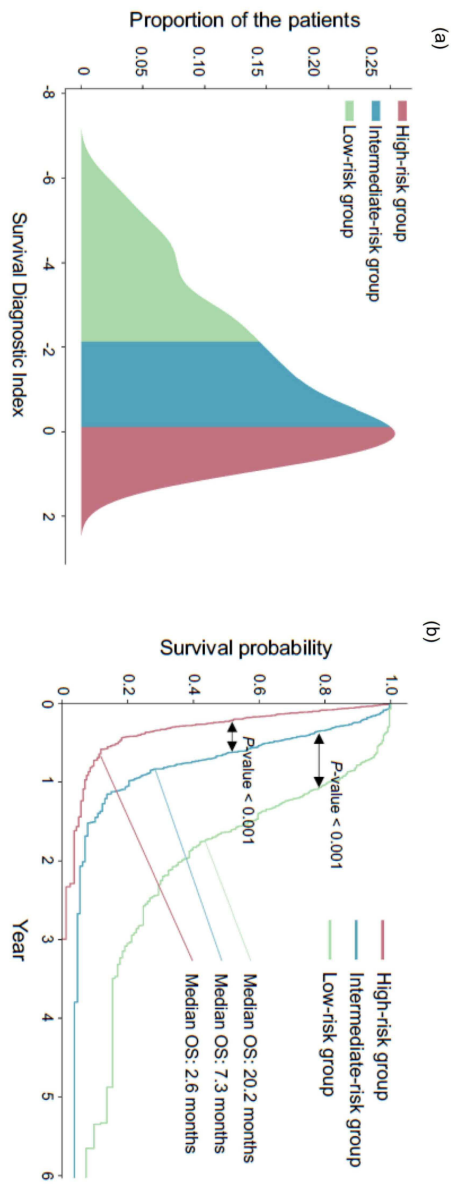
도면12



도면13



도면14



도면15

