

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0108682

(43) 공개일자 2021년09월03일

(51) 국제특허분류(Int. Cl.)

G16H 50/50 (2018.01) G06N 20/00 (2019.01)

G06N 3/08 (2006.01) G16H 30/20 (2018.01)

G16H 30/40 (2018.01) G16H 50/20 (2018.01)

G16H 50/30 (2018.01) G16H 50/70 (2018.01)

(52) CPC특허분류

G16H 50/50 (2018.01)

G06N 20/00 (2021.08)

(21) 출원번호 10-2020-0023584

(22) 출원일자 2020년02월26일

심사청구일자 2021년03월02일

(71) 출원인

계명대학교 산학협력단

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

(72) 발명자

송봉일

대구광역시 중구 동덕로 55, 102동 1803호(
대봉동, 대봉태왕아너스)

(74) 대리인

특허법인이룸리온

전체 청구항 수 : 총 9 항

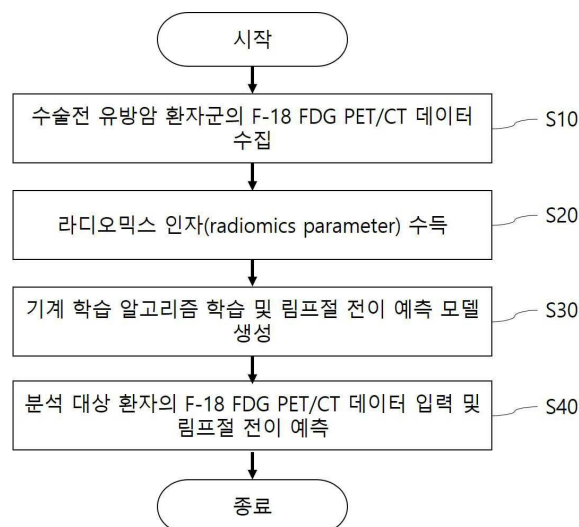
(54) 발명의 명칭 기계학습을 이용한 유방암 림프절 전이 예측에 대한 정보 제공 방법

(57) 요약

본 발명은 기계학습 알고리즘을 이용한 유방암 림프절 전이 예측에 대한 정보 제공 방법에 관한 것으로, 보다 상세하게는 유방암 환자의 F-18 FDG PET/CT의 원발암의 섭취 양상을 분석한 다음, 기계학습 분석을 수행하는 단계를 포함하는 유방암의 림프절 전이 예측에 대한 정보 제공 방법에 관한 것이다.

본 발명에서는 수술전 유방암 환자의 림프절 전이 예측을 위한 기계학습 모델을 확립하였으며, F-18 FDG PET/CT 소견과 XGBoost 알고리즘을 종합한 최종 림프절 전이 예측 결과는 90.9%의 진단 민감도, 71.4%의 진단 특이도, 80%의 진단 정확도를 보이는 것을 확인하였으므로, 본 발명은 유방암의 림프절 전이 여부에 대한 정보를 효과적으로 제공할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06N 3/08 (2013.01)
G16H 30/20 (2018.01)
G16H 30/40 (2018.01)
G16H 50/20 (2018.01)
G16H 50/30 (2018.01)
G16H 50/70 (2018.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	20190382
부처명	계명대학교
과제관리(전문)기관명	계명대학교
연구사업명	비사연구비
연구과제명	유방암에서 임파절 전이 예측을 위한 딥러닝 기반의 의료영상지표 모델
기 여 율	1/1
과제수행기관명	계명대학교
연구기간	2019.06.01 ~ 2020.05.31

명세서

청구범위

청구항 1

- (a) 복수의 기존 유방암 환자의 F-18 FDG PET/CT 원발암의 섭취 양상을 수집하는 단계;
- (b) 상기 수집된 데이터에 대한 텍스처 분석(texture analysis)을 수행하여 라디오믹스 인자(radiomics parameter)를 획득하는 단계;
- (c) 상기 (b) 단계에서 획득된 라디오믹스 인자를 기계학습 알고리즘으로 학습시켜 림프절 전이 예측 모델을 생성하는 단계;
- (d) 분석 대상 유방암 환자의 F-18 FDG PET/CT의 원발암의 섭취 양상을 상기 림프절 전이 예측 모델에 적용시키는 단계;를 포함하는 기계학습을 이용한 유방암의 림프절 전이 예측에 대한 정보 제공방법.

청구항 2

제1항에 있어서, 상기 (a) 단계는 유방암 환자에서 수술 전 시행한 F-18 FDG PET/CT 영상에서 원발 종양의 다이콤파일(dicom file)을 추출하고, 3D 슬라이서(3D slicer)에서 표준섭취계수(Standardized uptake value; SUV)가 2.5 이상인 부분을 관심 영역으로 분할(segmentation)을 시행하여 라벨이미지(label image)를 생성하는 것을 특징으로 하는 기계학습을 이용한 유방암의 림프절 전이 예측에 대한 정보 제공방법.

청구항 3

제1항에 있어서, 상기 (b) 단계의 텍스처 분석은 파이썬(python)의 파이라디오믹스(pyradiomics) 모듈을 사용하여 수행하는 것을 특징으로 하는 기계학습을 이용한 유방암의 림프절 전이 예측에 대한 정보 제공방법.

청구항 4

제1항에 있어서, 상기 (c) 단계는 엑스지부스트(XGBoost) 알고리즘을 이용하여 수행하는 것을 특징으로 하는 기계학습을 이용한 유방암의 림프절 전이 예측에 대한 정보 제공방법.

청구항 5

제1항에 있어서, 상기 (d) 단계는

- (i) 분석 대상 유방암 환자에서 수술 전 시행한 F-18 FDG PET/CT 영상에서 원발 종양의 다이콤파일(dicom file)을 추출하고, 3D 슬라이서(3D slicer)에서 표준섭취계수(Standardized uptake value; SUV)가 2.5 이상인 부분을 관심 영역으로 분할(segmentation)을 시행하여 라벨이미지(label image)를 생성하는 단계;
- (ii) 상기 다이콤파일(dicom file) 및 라벨이미지(label image)에 대한 텍스처 분석(texture analysis)을 수행하여 라디오믹스 인자(radiomics parameter)를 획득하는 단계;
- (iii) 획득된 라디오믹스 인자를 상기 (c) 단계의 림프절 전이 예측 모델에 적용시켜 림프절 전이 여부를 예측하는 단계;를 포함하는 방법으로 수행되는 것을 특징으로 하는 기계학습을 이용한 유방암의 림프절 전이 예측에 대한 정보 제공방법.

청구항 6

유방암 환자의 F-18 FDG PET/CT 영상 정보 입력부;

상기 F-18 FDG PET/CT 영상 정보를 처리하여 라디오믹스 인자를 획득하는 전처리부;

상기 라디오믹스 인자를 기계학습시켜 림프절 전이 모델을 생성하는 림프절 전이 모델 생성부;를 포함하는 기계 학습을 이용한 림프절의 전이 예측 장치.

청구항 7

제6항에 있어서, 상기 전처리부는

F-18 FDG PET/CT 영상에서 원발 종양의 다이코파일(dicom file)을 추출하는 다이코파일 추출부;

F-18 FDG PET/CT 영상의 3D 슬라이서(3D slicer)에서 표준섭취계수(Standardized uptake value; SUV)가 2.5 이상인 부분을 관심 영역으로 분할을 시행하여 라벨이미지(label image)를 생성하는 라벨이미지 생성부;

다이코파일 및 라벨이미지에 대한 텍스처 분석(texture analysis)을 수행하여 라디오믹스 인자(radiomics parameter)를 획득하는 라디오믹스 인자 획득부;를 포함하는 것을 특징으로 하는 기계학습을 이용한 림프절의 전이 예측 장치.

청구항 8

제6항에 있어서, 상기 림프절 전이 모델 생성부는 엑스지부스트(XGBoost) 알고리즘을 이용하여 림프절 전이 모델을 생성하는 것을 특징으로 하는 기계학습을 이용한 림프절의 전이 예측 장치.

청구항 9

제1항 내지 제5항 중 어느 한 항의 림프절 전이 예측에 대한 정보 제공방법을 실행하기 위해 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 기계학습을 이용한 유방암 림프절 전이 예측에 대한 정보 제공 방법에 관한 것으로, 보다 상세하게는 유방암 환자의 F-18 FDG PET/CT의 원발암의 섭취 양상을 분석하여 라디오믹스 인자(radiomics parameter)를 수득한 다음, 기계학습 분석을 수행하는 단계를 포함하는 유방암의 림프절 전이 예측에 대한 정보 제공 방법에 관한 것이다.

배경 기술

[0003] 유방암은 여성에게 발생하는 침습성 암 중 가장 흔한 암이며, 암관련 사망 원인의 두 번째를 차지한다. 임상적으로 종양의 크기, 종양의 조직학적 등급, 호르몬 수용체와 HER2, 액와림프절, 원격전이 등은 유방암 환자들의 관리 및 예후에 사용되어 왔다. 이 중 액와림프절 전이는 암의 치료 결정과 장기 예후를 결정하는데 가장 중요한 인자이다. 임상적으로 림프절 전이의 진단은 초음파, MRI, F-18 FDG PET/CT등을 이용하고 있지만 아직 낮은 민감도로 인해 감시림프절 절제술(sentinel lymph node biopsy)에 의존하고 있다.

[0004] 최근에는 다양한 암의 진단, 예후, 전이 예측의 민감도를 높이기 위해, 기계학습 알고리즘 분석을 통한 분석 방법이 활용되고 있다.

[0005] 이에, 본 발명에서는 유방암에서의 림프절 전이 예측의 민감도를 높일 수 있는 기계학습 기반의 의료영상지표 모델(radiomics model)을 개발하기 위해 예의 노력한 결과, 유방암 환자의 수술 전 시행한 F-18 FDG PET/CT 영상에서 원발 종양의 라디오믹스(radiomics) 파라미터들을 사용하여 수술전 유방암 환자의 림프절 전이 예측을

위한 기계학습 모델을 확립하였으며, F-18 FDG PET/CT 소견과 XGBoost 알고리즘을 종합한 최종 림프절 전이 예측 결과는 90.9%의 진단 민감도, 71.4%의 진단 특이도, 80%의 진단 정확도를 보이는 것을 확인하고, 본 발명을 완성하였다.

발명의 내용

해결하려는 과제

[0007] 본 발명의 목적은 기계학습 알고리즘을 이용한 유방암 림프절 전이 예측에 대한 정보 제공 방법 및 이를 이용한 유방암 림프절 전이 예측 장치를 제공하는 데 있다.

과제의 해결 수단

- [0009] 상술한 목적을 달성하기 위해,
- [0010] 본 발명은 (a) 복수의 기존 유방암 환자의 F-18 FDG PET/CT 원발암의 섭취 양상을 수집하는 단계;
- [0011] (b) 상기 수집된 데이터에 대한 텍스처 분석(texture analysis)을 수행하여 라디오믹스 인자(radiomics parameter)를 획득하는 단계;
- [0012] (c) 상기 (b) 단계에서 획득된 라디오믹스 인자를 기계학습 알고리즘으로 학습시켜 림프절 전이 예측 모델을 생성하는 단계;
- [0013] (d) 분석 대상 유방암 환자의 F-18 FDG PET/CT의 원발암의 섭취 양상을 상기 림프절 전이 예측 모델에 적용시키는 단계;
- [0014] 를 포함하는 기계학습을 이용한 유방암의 림프절 전이 예측에 대한 정보 제공방법을 제공한다.
- [0015] 본 발명의 바람직한 일실시예에 있어서, 상기 (a) 단계는 유방암 환자에서 수술 전 시행한 F-18 FDG PET/CT 영상에서 원발 종양의 다이콤파일(dicom file)을 추출하고, 3D 슬라이서(3D slicer)에서 표준섭취계수(Standardized uptake value; SUV)가 2.5 이상인 부분을 관심 영역으로 분할(segmentation)을 시행하여 라벨이미지(label image)를 생성하는 것일 수 있다.
- [0016] 본 발명의 바람직한 다른 일실시예에 있어서, 상기 (b) 단계의 텍스처 분석은 파이썬(python)의 파이라디오믹스(pyradiomics) 모듈을 사용하여 분석할 수 있다.
- [0017] 본 발명의 바람직한 또 다른 일실시예에 있어서, 상기 (c) 단계는 엑스지 부스트(XGBoost) 알고리즘을 이용하여 수행할 수 있다.
- [0018] 본 발명의 바람직한 또 다른 일실시예에 있어서, 상기 (d) 단계는 (i) 분석 대상 유방암 환자에서 수술 전 시행한 F-18 FDG PET/CT 영상에서 원발 종양의 다이콤파일(dicom file)을 추출하고, 3D 슬라이서(3D slicer)에서 표준섭취계수(Standardized uptake value; SUV)가 2.5 이상인 부분을 관심 영역으로 분할(segmentation)을 시행하여 라벨이미지(label image)를 생성하는 단계;
- [0019] (ii) 상기 다이콤파일(dicom file) 및 라벨이미지(label image)에 대한 텍스처 분석(texture analysis)을 수행하여 라디오믹스 인자(radiomics parameter)를 획득하는 단계;
- [0020] (iii) 획득된 라디오믹스 인자를 상기 (c) 단계의 림프절 전이 예측 모델에 적용시켜 림프절 전이 여부를 예측하는 단계;를 포함할 수 있다.
- [0022] 본 발명은 또한, 유방암 환자의 F-18 FDG PET/CT 영상 정보 입력부;
- [0023] 상기 F-18 FDG PET/CT 영상 정보를 처리하여 라디오믹스 인자를 획득하는 전처리부;
- [0024] 상기 라디오믹스 인자를 기계학습시켜 림프절 전이 모델을 생성하는 림프절 전이 모델 생성부;를 포함하는 기계학습을 이용한 림프절의 전이 예측 장치를 제공한다.

- [0025] 본 발명의 바람직한 일실시예에 있어서, 상기 전처리부는
- [0026] F-18 FDG PET/CT 영상에서 원발 종양의 다이콤파일(dicom file)을 추출하는 다이콤파일 추출부;
- [0027] F-18 FDG PET/CT 영상의 3D 슬라이서(3D slicer)에서 표준섭취계수(Standardized uptake value; SUV)가 2.5 이상인 부분을 관심 영역으로 분할을 시행하여 라벨이미지(label image)를 생성하는 라벨이미지 생성부;
- [0028] 다이콤파일 및 라벨이미지에 대한 텍스처 분석(texture analysis)을 수행하여 라디오믹스 인자(radiomics parameter)를 획득하는 라디오믹스 인자 획득부;로 구성될 수 있다.
- [0029] 본 발명의 바람직한 일실시예에 있어서, 상기 림프절 전이 모델 생성부는 엑스지 부스트(XGBoost) 알고리즘을 이용하여 림프절 전이 모델을 생성할 수 있다.
- [0031] 본 발명은 또한, 컴퓨터를 이용하여 상기 림프절 전이 예측에 대한 정보 제공방법을 실행하기 위해 매체에 저장된 컴퓨터 프로그램을 제공한다.

발명의 효과

- [0033] 본 발명에서는 수술전 유방암 환자의 림프절 전이 예측을 위한 기계학습 모델을 확립하였으며, F-18 FDG PET/CT 소견과 XGBoost 알고리즘을 종합한 최종 림프절 전이 예측 결과는 90.9%의 진단 민감도, 71.4%의 진단 특이도, 80%의 진단 정확도를 보이는 것을 확인하였으므로, 본 발명은 유방암의 림프절 전이 여부에 대한 정보를 효과적으로 제공할 수 있다.

도면의 간단한 설명

- [0035] 도 1은 본 발명의 일실시예에 따른 기계학습을 이용한 림프절 전이 예측 방법을 나타낸 순서도이다.
- 도 2는 3D slicer에서 표준섭취계수가 2.5 이상인 부분을 관심 영역으로 분할한 데이터이다.
- 도 3은 R statistics에서의 XGBoost를 이용한 기계학습 결과 분석된 림프절 전이 예측에 유용한 파라미터를 나타낸 데이터이다.
- 도 4는 XGBoost를 이용한 기계학습에서 림프절 전이 예측에 유용한 파라미터들의 중요도 순위를 나타낸 그래프이다. 도 4의 각 숫자가 의미하는 내용은 도 3과 동일하다.
- 도 5는 본 발명의 일실시예에 따른 기계학습을 이용한 림프절 전이 예측 장치의 구성을 개략적으로 나타낸 모식도이다.

발명을 실시하기 위한 구체적인 내용

- [0036] 이하, 본 발명을 상세하게 설명한다.
- [0038] 본 발명은 일관점에서, (a) 복수의 기존 유방암 환자의 F-18 FDG PET/CT 원발암의 섭취 양상을 수집하는 단계(S10);
- [0039] (b) 상기 수집된 데이터에 대한 텍스처 분석(texture analysis)을 수행하여 라디오믹스 인자(radiomics parameter)를 획득하는 단계(S20);
- [0040] (c) 상기 (b) 단계에서 획득된 라디오믹스 인자를 기계학습 알고리즘으로 학습시켜 림프절 전이 예측 모델을 생성하는 단계(S30);
- [0041] (d) 분석 대상 유방암 환자의 F-18 FDG PET/CT의 원발암의 섭취 양상을 상기 림프절 전이 예측 모델에 적용시키는 단계(S40);를 포함하는 기계학습을 이용한 유방암의 림프절 전이 예측에 대한 정보 제공방법에 관한 것이다.

- [0043] 도 1은 본 발명의 일실시예에 따른 기계학습을 이용한 림프절 전이 예측 방법을 나타낸 순서도이다. 도 1에서 S10은 수술전 유방암 환자군의 F-18 FDG PET/CT 데이터 수집하는 단계로, 본 발명의 구체적인 일실시예에서는, 2007년 12월부터 20011년 12월까지 유방암 환자의 수술 전 시행한 F-18 FDG PET/CT 영상에서 원발 종양의 다이콤파일(dicom file)을 추출하고, 3D 슬라이서(3D slicer)에서 표준섭취계수(Standardized uptake value; SUV)가 2.5 이상인 부분을 관심 영역으로 분할(segmentation)을 시행하여 라벨이미지(label image)를 생성하였다(도 2).
- [0044] 도 1의 S20은 상기 S10에서 수집된 데이터에 대한 텍스처 분석(texture analysis)을 수행하여 라디오믹스 인자(radiomics parameter)를 획득하는 단계이다. 본 발명의 구체적인 일실시예에서 72명의 환자의 F-18 FDG PET/CT 영상에서 획득한 다이콤파일(dicom file) 및 라벨이미지(label image)를 파이썬(python)의 파이라디오믹스(pyradiomics) 모듈을 사용하여 텍스처 분석(texture analysis)을 수행하여 라디오믹스 인자(radiomics parameter)를 획득하였다. 각 환자별로 총 793개의 라디오믹스 인자들을 추출하였다.
- [0045] 도 1의 S30은 S20에서 획득된 라디오믹스 인자를 기계학습 알고리즘으로 학습시켜 림프절 전이 예측 모델을 생성하는 단계이다. 본 발명의 구체적인 일실시예에서, 각 환자별로 추출한 라디오믹스 인자들을 R 스타티스(R statistics)에서 XGBoost 모듈을 사용하여 기계학습을 수행하였으며, 도 3의 파라미터들이 림프절 전이를 예측하는데 유용한 파라미터들로 분석되었고, 이를 중요도에 따른 그래프로 나타내었다(도 4).
- [0046] 도 1의 S40은 분석 대상 유방암 환자의 F-18 FDG PET/CT 데이터를 상기 림프절 전이 예측 모델에 입력하여 림프절 전이를 예측하는 단계로, 상기 F-18 FDG PET/CT 데이터는 S10 및 S20과 동일하게 처리하여 라디오믹스 인자를 획득한 다음, 이를 림프절 전이 예측 모델에 적용시켜 최종적으로 림프절 전이 여부를 예측하게 된다.
- [0047] 본 발명의 구체적인 일실시예에서, XGBoost 알고리즘으로 학습 세트(training set)에서 도출된 림프절 전이 예측 모델에 총 25명의 검증 세트(test set) 환자 중 F-18 FDG PET/CT에서 림프절 전이 소견을 보이지 않은 16명의 환자를 적용시킨 결과, 11명은 림프절 전이가 없는 것으로 예측되었다. 이 환자들 중 10명은 실제 림프절 전이가 없었고, 한명은 실제 림프절 전이가 있었으나 위음성으로 예측되었다. 5명은 림프절 전이가 있는 것으로 예측되었는데, 이 환자들 중 2명은 실제 림프절 전이가 있었고, 3명은 위양성으로 예측되었다.
- [0048] 총 25명의 검증 세트에서 F-18 FDG PET/CT 소견과 XGBoost 알고리즘을 종합한 최종 림프절 전이 예측 결과는 90.9%의 진단 민감도, 71.4%의 진단 특이도, 80%의 진단 정확도를 보였으며, 25명의 검증에서 F-18 FDG PET/CT 소견만으로 얻은 72.7%의 진단 민감도, 92.8%의 진단 특이도, 84%의 진단 정확도와 비교하여 진단 특이도와 진단 정확도는 다소 감소하였으나, 진단 민감도가 현저하게 향상된 것을 확인하였다.
- [0050] 본 발명은 다른 일관점에서,
- [0051] 유방암 환자의 F-18 FDG PET/CT 영상 정보 입력부(11);
- [0052] 상기 F-18 FDG PET/CT 영상 정보를 처리하여 라디오믹스 인자를 획득하는 전처리부(12);
- [0053] 상기 라디오믹스 인자를 기계학습시켜 림프절 전이 모델을 생성하는 림프절 전이 모델 생성부(16);를 포함하는 기계학습을 이용한 림프절의 전이 예측 장치(10)를 제공한다.
- [0054] 도 5는 본 발명의 일실시예에 따른 기계학습을 이용한 림프절 전이 예측 장치의 구성을 개략적으로 나타낸 모식도이다.
- [0055] 도 5에 도시된 림프절 전이 예측 장치(10)는 본 실시예의 특징이 흐려지는 것을 방지하기 위하여 본 실시예와 관련된 구성요소들만을 도시한 것이다. 따라서, 도 5에 도시된 구성요소들 외에 다른 범용적인 구성요소들이 더 포함될 수 있음을 본 실시예와 관련된 기술분야에서 통상의 지식을 가진 자라면 이해할 수 있다.
- [0056] 본 발명의 일 실시예에 따른 림프절 전이 예측 장치(10)는 적어도 하나 이상의 프로세서(processor)에 해당하거나, 적어도 하나 이상의 프로세서를 포함할 수 있다. 이에 따라, 림프절 전이 예측 장치(10)는 마이크로프로세서나 범용 컴퓨터 시스템과 같은 다른 하드웨어 장치에 포함된 형태로 구동될 수 있다.
- [0057] 도 5를 참조하면, 림프절 전이 예측 장치(10)는 입력부(11), 전처리부(12) 및 림프절 전이 모델 생성부(16)를 포함하며, 전처리부(12)는 다이콤파일 추출부(13), 라벨이미지 생성부(14) 및 라디오믹스인자 획득부(15)를 포함한다.

- [0058] 입력부(11)는 복수의 수술전 유방암 환자 F-18 FDG PET/CT 영상 데이터를 입력한다.
- [0059] 전처리부(12)는 구체적으로, 상기 전처리부는 F-18 FDG PET/CT 영상에서 원발 종양의 다이콤파일(dicom file)을 추출하는 다이콤파일 추출부(13); F-18 FDG PET/CT 영상의 3D 슬라이서(3D slicer)에서 표준섭취계수(Standardized uptake value; SUV)가 2.5 이상인 부분을 관심 영역으로 분할을 시행하여 라벨이미지(label image)를 생성하는 라벨이미지 생성부(14); 다이콤파일 및 라벨이미지에 대한 텍스처 분석(texture analysis)을 수행하여 라디오믹스 인자(radiomics parameter)를 획득하는 라디오믹스 인자 획득부(15);로 구성된다.
- [0060] 림프절 전이 모델 생성부(16)는 전처리부(12)의 라디오믹스 인자 획득부(15)에서 획득한 라디오믹스 인자를 기계학습시켜 림프절 전이 모델을 생성하는 것으로, 엑스지 부스트(XGBoost) 알고리즘을 이용하여 림프절 전이 모델을 생성할 수 있다. 이 외에도, 기계학습에는 인공신경망(artificial neural network) 등 다양한 알고리즘이 활용될 수 있다.
- [0061] 또한, 상기 림프절 전이 모델 생성부(16)는 림프절 전이 예측에 유용한 라디오믹스 인자에 중요도 순위를 부여하고, 상기 중요도 순위를 기반으로 최종 입력 라디오믹스 인자를 선택하는 처리부가 포함될 수 있다.
- [0063] 본 발명은 또 다른 관점에서, 컴퓨터를 이용하여 상기 림프절 전이 예측에 대한 정보 제공방법을 실행하기 위해 매체에 저장된 컴퓨터 프로그램에 관한 것이다.
- [0064] 도 1에 도시된 본 발명의 일실시예에 따른 기계학습을 이용한 유방암의 림프절 전이 예측에 대한 정보 제공방법은 컴퓨터에서 실행될 수 있는 프로그램으로 작성할 수 있고, 컴퓨터로 읽을 수 있는 기록매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다. 상기 컴퓨터로 읽을 수 있는 기록매체는 마그네틱 저장매체(예를 들면, 롬, 플로피 디스크, 하드 디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등)와 같은저장매체를 포함한다.
- [0065] 본 발명에 따른 기계학습을 이용한 유방암의 림프절 전이 예측에 대한 정보 제공방법, 장치 및 프로그램에 따르면, 유방암 환자의 림프절 전이 여부를 개개인별로 정확하게 예측하여 환자별 맞춤형 치료 방법을 결정할 수 있다.
- [0067] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다.
- [0068] 이들 실시예는 오로지 본 발명을 예시하기 위한 것으로서, 본 발명의 범위가 이들 실시예에 의해 제한되는 것으로 해석되지 않는 것은 당업계에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.

실시예 1

[0070] 환자 분류 및 임상 정보

- [0071] 2007년 12월부터 2011년 12월까지 계명대학교 동산의료원에서 유방암으로 외과적 치료를 받은 133명의환자들의 의무 기록을 후향적(retrospective)으로 조사하였다. 이 중 원발암의 F-18 FDG가 없거나 미약하여 원발 종양의 라디오믹스 인자(radiomics parameter)들을 분석하기 불가능한 환자는 33명이었고, 이를 제외한 100명의 환자를 본 연구의 분석에 사용하였다.
- [0072] 전체 100명의 환자 중 림프절 전이가 있는 환자는 43명 (43%) 이었다. F-18 FDG PET/CT에서 림프절에 이상 섭취가 관찰된 환자는 총 28명 (28%) 이었고, 이 환자들 중 실제 림프절 전이가 병리학적으로 관찰된 환자는 24명으로 F-18 FDG PET/CT의 유방암에서 림프절 전이 진단 민감도는 55.8%, 진단 특이도는 93%, 진단 정확도는 77%이었다.

실시예 2

[0074] 기계학습을 이용한 림프절 전이 예측 모델 확립

- [0075] 기계학습을 이용한 림프절 전이 예측 모델 확립하기 위해, 전체 100명의 환자군을 75명의 학습 세트(training

set)와 25명의 검증 세트(test set)로 무작위로 나누었고, F-18 FDG PET/CT에서 림프절 전이 소견을 보이지 않는 72명의 환자는 학습 세트에는 56명이, 검증 세트에는 16명이 분포하는 것으로 확인되었다.

[0076] 72명의 환자의 F-18 FDG PET/CT 영상에서 원발 종양의 다이콤파일(dicom file)을 추출하고, 3D 슬라이서(3D slicer)에서 표준섭취계수(Standardized uptake value; SUV)가 2.5 이상인 부분을 관심 영역으로 분할(segmentation)을 시행하여 라벨이미지(label image)를 생성하였다 (도 2).

[0077] 수득한 다이콤파일(dicom file) 및 라벨이미지(label image)를 파이썬(python)의 파이라디오믹스(pyradiomics) 모듈을 사용하여 텍스처 분석(texture analysis)을 수행하여 라디오믹스 인자(radiomics parameter)를 수득하였다. 각 환자별로 총 793개의 라디오믹스 인자들을 추출하였다.

[0078] 이후, 추출한 라디오믹스 인자들을 R 스태틱스(R statistics)에서 XGBoost 모듈을 사용하여 기계학습을 수행하고, 림프절 전이 예측 모델을 확립하였다.

[0079] 그 결과, 도 3에 나타난 바와 같이, 림프절 전이를 예측하는데 유용한 파라미터들이 분석되었으며, 이를 중요도에 따른 그래프로 나타내었다 (도 4).

[0081] XGBoost 알고리즘으로 학습 세트(training set)에서 도출된 림프절 전이 예측 모델에 F-18 FDG PET/CT에서 림프절 전이 소견을 보이지 않은 16명의 검증 세트(test set)를 적용시킨 결과, 11명은 림프절 전이가 없는 것으로 예측되었다. 이 환자들 중 10명은 실제 림프절 전이가 없었고, 한명은 실제 림프절 전이가 있었으나 위음성으로 예측되었다. 5명은 림프절 전이가 있는 것으로 예측되었는데, 이 환자들 중 2명은 실제 림프절 전이가 있었고, 3명은 위양성으로 예측되었다.

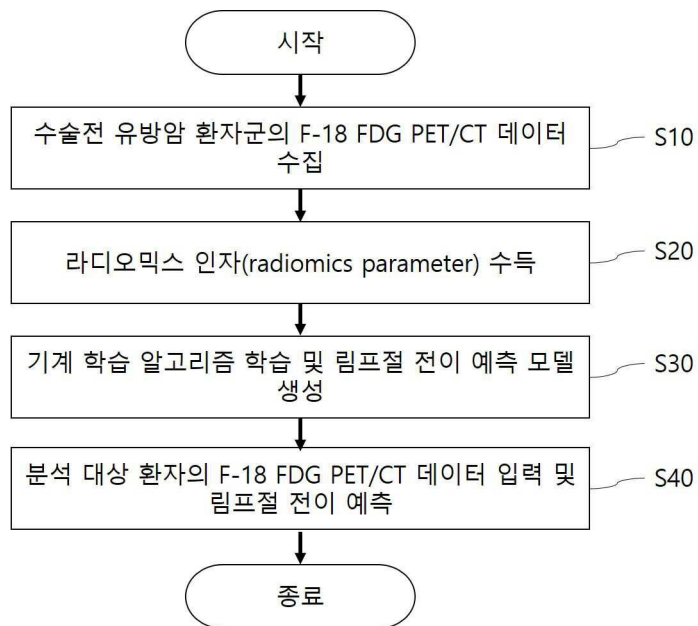
[0082] 총 25명의 검증 세트에서 F-18 FDG PET/CT 소견과 XGBoost 알고리즘을 종합한 최종 림프절 전이 예측 결과는 90.9%의 진단 민감도, 71.4%의 진단 특이도, 80%의 진단 정확도를 보였으며, 25명의 검증에서 F-18 FDG PET/CT 소견만으로 얻은 72.7%의 진단 민감도, 92.8%의 진단 특이도, 84%의 진단 정확도와 비교하여 진단 특이도와 진단 정확도는 다소 감소하였으나, 진단 민감도가 현저하게 향상된 것을 확인하였다. 임상적으로는, F-18 FDG PET/CT 소견과 XGBoost 알고리즘을 종합한 최종 림프절 전이 예측 결과에서 림프절 전이가 없는 것으로 나오는 환자의 경우, 감시림프절 절제술을 생략하고 수술적 치료를 시행할 수 있는데 그 의의가 있다고 할 수 있다.

부호의 설명

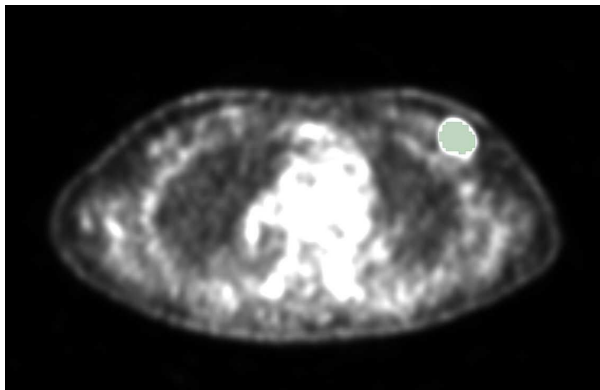
- [0084]
- 11 : 입력부
 - 12 : 전처리부
 - 13 : 다이콤파일 추출부
 - 14 : 라벨이미지 생성부
 - 15 : 라디오믹스 인자 획득부
 - 16 : 림프절 전이 모델 생성부

도면

도면1



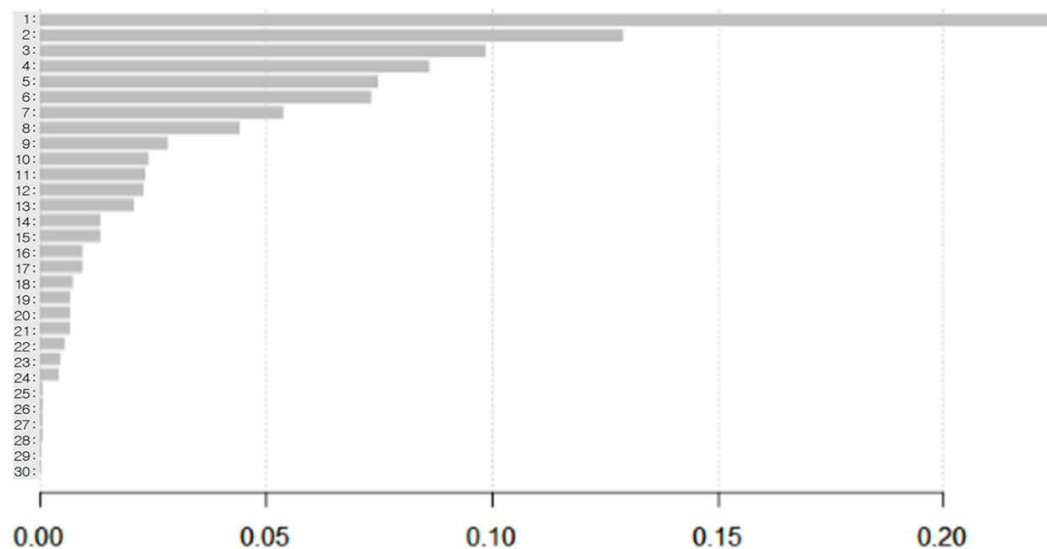
도면2



도면3

	Feature	Gain	Cover	Frequency
1:	wavelet.LHH_glszm_GrayLevelNonUniformityNormalized	0.2247461291	0.293846039	0.209174546
2:	wavelet.LLL_glcm_JointEnergy	0.1292586115	0.137967886	0.099412216
3:	original_shape_Sphericity	0.0986007917	0.057702044	0.087656530
4:	wavelet.LHL_glcm_ClusterTendency	0.0863318219	0.123677958	0.083056478
5:	wavelet.LLL_glrIm_LongRunLowGrayLevelEmphasis	0.0748353196	0.068062622	0.046256070
6:	wavelet.LHL_glcm_Contrast	0.0736520457	0.017651398	0.046256070
7:	wavelet.LLH_glcm_Correlation	0.0539546860	0.033761228	0.041656018
8:	wavelet.LHH_glrIm_LongRunEmphasis	0.0443721632	0.018727049	0.035778175
9:	wavelet.HHL_firstorder_Minimum	0.0285690287	0.042885361	0.035778175
10:	wavelet.HHL_glrIm_GrayLevelNonUniformityNormalized	0.0241310697	0.014353581	0.024533606
11:	wavelet.LLL_glrIm_ShortRunLowGrayLevelEmphasis	0.0233796125	0.020276218	0.027344748
12:	wavelet.LHH_glszm_SmallAreaEmphasis	0.0233020845	0.016853648	0.028878099
13:	wavelet.HLH_glcm_MaximumProbability	0.0211308635	0.011747057	0.027344748
14:	wavelet.HHH_glszm_GrayLevelNonUniformityNormalized	0.0137045802	0.019990998	0.027728086
15:	wavelet.HHL_glszm_SmallAreaHighGrayLevelEmphasis	0.0136639968	0.019918678	0.027600307
16:	wavelet.HLL_glrIm_ShortRunHighGrayLevelEmphasis	0.0097529214	0.018428657	0.019422438
17:	wavelet.LHH_glrIm_HighGrayLevelRunEmphasis	0.0097293283	0.009739769	0.014055712
18:	wavelet.LHH_glcm_Lmc1	0.0074170174	0.009977633	0.016483516
19:	wavelet.LHH_glcm_Contrast	0.0068966702	0.010310212	0.009200102
20:	original_gldm_LargeDependenceHighGrayLevelEmphasis	0.0067542935	0.005773388	0.014055712
21:	wavelet.HHH_glcm_Correlation	0.0066761171	0.012774226	0.012777920
22:	wavelet.LLL_glszm_SmallAreaEmphasis	0.0056670299	0.004719477	0.014822387
23:	wavelet.LHH_glrIm_LongRunHighGrayLevelEmphasis	0.0047729486	0.007058153	0.006261181
24:	wavelet.HHL_glszm_SizeZoneNonUniformityNormalized	0.0043986694	0.003809683	0.012522361
25:	wavelet.HHH_glszm_SmallAreaEmphasis	0.0009349452	0.005218724	0.007666752
26:	wavelet.HHH_glrIm_GrayLevelNonUniformityNormalized	0.0009348629	0.005218811	0.007666752
27:	wavelet.HHH_glszm_SmallAreaHighGrayLevelEmphasis	0.0009348583	0.005218937	0.007666752
28:	original_firstorder_InterquartileRange	0.0008958204	0.001047664	0.003833376
29:	wavelet.LLH_glrIm_ShortRunHighGrayLevelEmphasis	0.0003156856	0.001720868	0.002683363
30:	wavelet.HLL_firstorder_Mean	0.0002860272	0.001562033	0.002427805

도면4



도면5

