



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년01월06일
(11) 등록번호 10-2485316
(24) 등록일자 2023년01월02일

- (51) 국제특허분류(Int. Cl.)
G16B 40/20 (2019.01) C12Q 1/6886 (2018.01)
G06N 20/00 (2019.01) G06N 3/08 (2023.01)
G16B 20/20 (2019.01) G16B 20/50 (2019.01)

- (52) CPC특허분류
G16B 40/20 (2019.02)
C12Q 1/6886 (2022.01)

- (21) 출원번호 10-2022-0016286

- (22) 출원일자 2022년02월08일

- 심사청구일자 2022년02월08일

- (56) 선행기술조사문헌

H. Gunasekaran 외, "Analysis of DNA sequence classification using CNN and hybrid models", Computational and Mathematical Methods in Medicine, 2021.*

S. Kwon 외, "Personal health information inference using machine learning on RNA expression data from patients with cancer: Algorithm validation study", J. of Medical Internet Research, 22권, 8호, 2020.*

Y. Sun 외, "Identification of 12 cancer types through genome deep learning", Scientific Reports, 9:17256, 2019.*

*는 심사관에 의하여 인용된 문헌

- (73) 특허권자

한밭대학교 산학협력단

대전광역시 유성구 동서대로 125 (덕명동)

주식회사 쓰리빅스

부산광역시 서구 구덕로 187, 융합의학연구동 3층
유알피 3-3호 (아미동1가)

- (72) 발명자

이승호

대전광역시 유성구 엑스포로123번길 26-30 스마트
시티리버뷰오피스텔 102동 601호

이영지

대전광역시 대덕구 계족로761번길 24 현대아파트
114동 707호

박준형

경기도 화성시 동탄지성로 42 동탄시범한빛마을
동탄아이파크 224동 404호

- (74) 대리인

이은철

전체 청구항 수 : 총 5 항

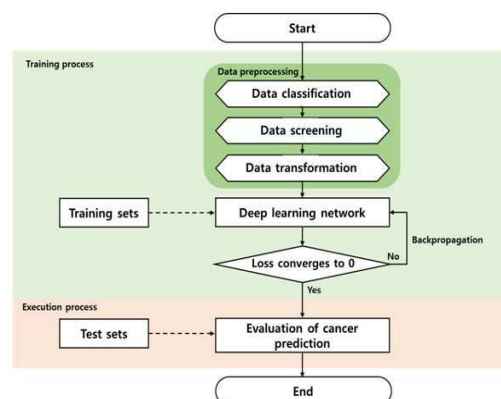
심사관 : 성경아

(54) 발명의 명칭 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법

(57) 요약

본 발명은 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측을 위한 방법에 관한 것으로, 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법에 관한 것이다. 본 발명에 따르면, 데이터 분류, 데이터 선별, 데이터 변환 등을 처리하는 데이터 전처리 부분과 데이터를 학습하는 딥러닝 구조를 수행하여 암 발생과의 연관성을 예측하는 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법을 제공하여, 특정 암에 관련된 유전체 패턴을 분석하지 않고 전체적인 유전체 패턴을 분석하기 위해 개인별 전체 유전체 데이터를 입력으로 사용하므로, 개인에 대한 여러 암에 대해 예측 활용도가 높은 효과가 있다.

대표도 - 도3



(52) CPC특허분류

G06N 20/00 (2021.08)

G06N 3/08 (2023.01)

G16B 20/20 (2019.02)

G16B 20/50 (2019.02)

C12Q 2600/118 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1425138116

과제번호 S2788640

부처명 중소벤처기업부

과제관리(전문)기관명 중소기업기술정보진흥원

연구사업명 창업성장기술개발(R&D)

연구과제명 건강 및 질병관리를 위한 Multi-Omics를 활용한 유전체 분석 플랫폼 개발

기 여 율 1/1

과제수행기관명 주식회사 쓰리빅스

연구기간 2019.10.01 ~ 2021.09.30

공지예외적용 : 있음

명세서

청구범위

청구항 1

(a)개인별 데이터로 분류하는 데이터 분류과정, 학습 결과에 대한 정확도와 학습 시간을 줄여주는 데이터 선별 과정, 학습이 가능한 형태로 변환하는 데이터 변환과정을 포함하는 데이터 전처리 단계 및

(b)전처리한 데이터를 학습하는 딥러닝 구조를 수행하여 암 발생과의 연관성을 예측하는 단계를 포함하되,

개인의 전체유전체 데이터로 분류한 후에 학습데이터에 사용할 변수 정보를 선별하고, 선별된 유전체 데이터 변수 정보를 기반으로 학습데이터를 구축하며,

상기 (a) 단계는 개인별 전체 유전체 데이터를 입력으로 하여 각 생체 샘플 데이터들을 구분하기 위한 고유키인 Sample_Barcode를 사용하여 개인별 전체 유전체 데이터로 분류하고, 상기 개인별 전체 유전체 데이터로 분류한 후에, 데이터 선별과정에서 환경적 요인 및 후천적인 요인을 학습데이터에 사용할 변수 정보에서 제외하는 것을 특징으로 하는 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법.

청구항 2

제1항에 있어서,

상기 (a)단계는 데이터 분류 및 선별 과정 후, 개인의 유전체 데이터를 하나의 문장으로 변환하는 것을 특징으로 하는 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법.

청구항 3

제1항에 있어서,

상기 (a)단계는 개인의 여러 유전체 데이터를 하나의 문장으로 변환한 후에 대문자를 소문자로 바꾸고 one hot 인코딩을 수행하는 것을 특징으로 하는 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법.

청구항 4

제1항에 있어서,

상기 (b)단계는 입력 데이터에서 다양한 특징들을 추출하는 과정, 데이터의 일정 구간마다 최대값을 추출하여 다운 샘플링하는 제1 폴링과정, 학습데이터의 각 row에서 가장 값이 높은 k개를 추출하여 다운 샘플링을 수행하는 제2 폴링과정을 수행하고, 마지막단에 뉴런의 출력값에 대하여 클래스를 분류하기 위해 각 클래스에 대한 확률을 추정하는 과정을 포함하는 것을 특징으로 하는 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법.

청구항 5

제1항에 있어서,

상기 학습데이터에 사용되는 변수 정보는 염색체 번호, 염색체 위치, 유전자 변이, 변이 종류, 성별을 포함하는 것을 특징으로 하는 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법.

발명의 설명

기술 분야

본 발명은 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측을 위한 방법에 관한 것으로, 보다 상세하게는 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법에 관한 것이다.

[0001]

배경 기술

- [0002] 현재 암 진단을 받는 환자 수는 계속해서 증가하고 있으며, WHO는 지난 20년간 암 진단을 받은 전체 인구수는 2000년 1,000만 명에서 2020년 1,930만 명으로 거의 2배 정도 증가했다고 발표하였다. 2020년 한 해 동안만 전 세계적으로 약 1,930만 명의 암 환자가 발생했으며, 1,000만 명 정도가 암으로 사망하였다고 보도되었다. 또한, 인구 고령화로 인해 전체 사망자 중 암 사망자의 비율이 지속적으로 증가하고 있다. 도 1과 같이 2040년에는 전 세계적으로 암 환자가 약 2,843만 명으로 증가하여 2020년보다 1.5배 정도 증가할 것으로 예상된다.
- [0003] 암을 예측하는 것은 의료비 절감에도 상당히 중요한 역할을 하며, 암을 초기에 예방하는 것은 생존율을 높여준다. 이러한 방안으로 개인의 건강 상태를 확인하는 바이오 헬스케어 시장규모가 세계적으로 증가하고 있으며, 의료 서비스 질을 높이고 비용을 줄이는 것이 최근 추세이다.
- [0004] 한편, 암을 발생시키는 원인이 많지만, 유전자 및 염색체 변이가 발생할 때 암이 발생할 확률이 매우 높다고 알려져 있다. 유전자의 역할은 인체를 자연스럽게 성장과 발달시키고 생명을 유지한다. 그러나 유전자의 아주 작은 부분이 결실되거나 이상이 생기면 이로 인해 각종 기형, 질환, 심한 경우 사망에 이르기까지 다양한 결과를 초래한다. 의학의 발전으로 유전자들이 질병과 어떤 연관성이 있는지에 대한 상당 부분이 밝혀지고 있으며, 돌연변이 유전자, 특정 질환과 연관된 유전자 변이 등의 연구가 계속되고 있다. 암은 대부분 유전자 돌연변이로 발생하며, 흡연, 자외선, 환경 변화 등으로 인해 유전자 돌연변이가 가장 일반적인 암 발생 패턴이다.
- [0005] 연구에 의해 밝혀진 대표적인 돌연변이에는 유방암 발생 위험이 있는 BRCA1, BRCA2 등 돌연변이 유전자가 있다. 이 유전자에 돌연변이가 있는 경우, 여성은 유방암 발생 위험이 60 ~ 80%이며, 난소암 발생 위험이 40% 정도 있다. 남성의 경우, 전립선암과 남성 유방암의 위험이 커지고 췌장암, 대장암 등과의 관련이 있다고 보고되고 있다. 이처럼 유전자와 질병과의 연관성에 관한 연구는 매우 중요하고 꾸준히 진행되고 있다. 돌연변이는 발생 원인에 따라 고유한 특징을 가지므로 특정 소수 유전자의 패턴 분석뿐만 아니라 전체적인 유전자의 패턴 분석이 요구되고 있지만 수천 개의 유전자 패턴을 분석하기에는 많은 어려움이 있다. 따라서 최근에는 딥러닝을 이용하여 다양하고 복잡한 데이터를 학습시켜 결과를 추출하는 연구가 진행되고 있다.

선행기술문헌

특허문헌

- [0006] (특허문헌 0001) 등록특허 10-2044094 (등록일 2019.11.06.)

발명의 내용

해결하려는 과제

- [0007] 본 발명은 상술한 문제를 해결하고자 고안한 것으로, 데이터 분류, 데이터 선별, 데이터 변환 등을 처리하는 데이터 전처리 부분과 데이터를 학습하는 딥러닝 구조를 수행하여 암 발생과의 연관성을 예측하는 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법을 제공함에 목적이 있다.

과제의 해결 수단

- [0008] 본 발명에 따른 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법은 (a)개인 별 데이터로 분류하는 데이터 분류과정, 학습 결과에 대한 정확도와 학습 시간을 줄여주는 데이터 선별과정, 학습이 가능한 형태로 변환하는 데이터 변환과정을 포함하는 데이터 전처리 단계 및 (b)전처리한 데이터를 학습하는 딥러닝 구조를 수행하여 암 발생과의 연관성을 예측하는 단계를 포함하되, 개인의 전체유전체 데이터로 분류한 후에 학습데이터에 사용할 변수 정보를 선별하고, 선별된 유전체 데이터 변수 정보를 기반으로 학습데이터를 구축한다.
- [0009] 상기 (a)단계는 데이터 분류 및 선별 과정 후, 개인의 유전체 데이터를 하나의 문장으로 변환한다.
- [0010] 상기 (a)단계는 개인의 여러 유전체 데이터를 하나의 문장으로 변환한 후에 대문자를 소문자로 바꾸고 one hot 인코딩을 수행한다.
- [0011] 상기 (b)단계는 입력 데이터에서 다양한 특징들을 추출하는 과정, 데이터의 일정 구간마다 최대값을 추출하여

다운 샘플링하는 제1 풀링과정, 학습데이터의 각 row에서 가장 값이 높은 k개를 추출하여 다운 샘플링을 수행하는 제2 풀링과정을 수행하고, 마지막단에 뉴런의 출력값에 대하여 클래스를 분류하기 위해 각 클래스에 대한 확률을 추정하는 과정을 포함한다.

[0012] 상기 학습데이터에 사용되는 변수 정보는 염색체 번호, 염색체 위치, 유전자 변이, 변이 종류, 성별을 포함한다.

발명의 효과

[0013] 본 발명의 일 실시예에 따르면, 특정 암에 관련된 유전체 패턴을 분석하지 않고 전체적인 유전체 패턴을 분석하기 위해 개인별 전체 유전체 데이터를 입력으로 사용하므로, 개인에 대한 여러 암에 대해 예측 활용도가 높다.

[0014] 개인의 전체적인 유전체 데이터를 입력으로 사용함으로써 정상인에 대한 참조 유전체 서열 데이터가 별도로 필요하지 않다. 특정 질환의 유전체 변이패턴뿐만 아니라 전체적인 유전체 패턴을 분석할 수 있다. 유전체 데이터를 사용할 때, 유전자 염기쌍 하나가 치환, 삽입, 결실이 발생하는 점 돌연변이뿐만 아니라 여러 염기쌍의 치환, 삽입, 결실의 일반적인 변이도 포함한다. Character convolution network 기반의 딥러닝 네트워크를 사용하여 학습을 수행하므로 다양한 질환, 희귀질환 등에도 쉽게 적용할 수 있다.

도면의 간단한 설명

[0015] 도 1은 2020년부터 2040년까지의 암환자 증가율을 나타낸 그래프이다.

도 2는 본 발명의 일 실시예에 따른 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측의 전체 개요도이다.

도 3은 본 발명의 일 실시예에 따른 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법의 흐름도이다.

도 4는 TCGA 데이터를 Sample_Barcode를 사용하여 개인별 전체 유전체 데이터로 분류한 것을 나타낸다.

도 5는 유전체 데이터에서 제외할 변수 정보를 나타낸 도면이다.

도 6은 선별된 유전체 데이터의 정보를 나타낸 도면이다.

도 7은 학습데이터에 사용할 변수 정보를 나타내고 있다.

도 8은 개인의 유전체 데이터를 하나의 문장으로 변환하는 과정을 나타낸 도면이다.

도 9는 one hot 인코딩을 수행한 예시를 나타낸 도면이다.

도 10은 개인 유전체 데이터에 대한 데이터 변환을 나타낸 도면이다.

도 11은 12가지 암에 대한 개인 유전체 학습 데이터 구축을 설명하기 위한 도면이다.

도 12는 딥러닝 네트워크 구조 설계를 설명하기 위한 도면이다.

도 13은 Max Pooling을 설명하기 위한 도면이다.

도 14는 Max pooling과 K-max Pooling을 비교하기 위한 도면이다.

도 15는 Fully connected layer를 설명하기 위한 FC layer이다.

도 16은 활성화 함수를 나타낸 그래프이다.

도 17은 마지막 단에 Softmax를 사용한 이유를 설명하기 위한 도면이다.

도 18은 본 발명의 일 실시예에 따른 backpropagation을 적용한 개요도이다.

도 19는 모델의 성능을 평가하는 방법을 나타내고 있다.

도 20은 실제 암에 걸린 수치와 암을 예측한 수치를 나타낸 도면이다.

도 21은 본 발명에서 예측한 12가지 암에 대한 ROC curve를 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0016] 본 발명의 실시예에서 제시되는 특정한 구조 내지 기능적 설명들은 단지 본 발명의 개념에 따른 실시예를 설명하기 위한 목적으로 예시된 것으로, 본 발명의 개념에 따른 실시예들은 다양한 형태로 실시될 수 있다. 또한, 본 명세서에 설명된 실시예들에 한정되는 것으로 해석되어서는 아니 되며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경물, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0017] 한편, 본 발명에서 제1 및/또는 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 한정되지는 않는다. 상기 용어들은 하나의 구성요소를 다른 구성요소들과 구별하는 목적으로만, 예컨대 본 발명의 개념에 따른 권리 범위로부터 벗어나지 않는 범위 내에서, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.
- [0018] 이하 첨부된 도면을 참조하여 본 발명의 실시예를 설명한다. 본 발명의 실시예를 설명함에 있어서, 관련된 공지 기능 혹은 구성에 대한 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 설명을 생략하였다.
- [0019] 도 2는 본 발명의 일 실시예에 따른 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법의 전체 개요도이다.
- [0020] 본 발명에서 제안하는 기법의 개요는 도 2과 같다. 제안된 기법은 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측시스템의 데이터 전처리부(110), 딥러닝 네트워크 구조부(120) 구성을 포함한다.
- [0021] 데이터 전처리부(110)는 데이터 분류, 데이터 선별, 데이터 변환 등을 처리하는 구성이다.
- [0022] 딥러닝 네트워크 구조부(120)는 전처리한 데이터를 딥러닝 네트워크 구조로 학습하기 위한 구성이다.
- [0023] 본 발명의 전체 개요도는 도 2와 같다. 개인별 전체 유전체 데이터와 암 발생과의 연관성을 딥러닝 네트워크를 통해 예측한다.
- [0024] 도 3은 본 발명의 일 실시예에 따른 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 방법의 흐름도이다. 본 발명의 흐름도는 데이터 분류, 데이터 선별, 데이터 변환 등을 처리하는 데이터 전처리 부분과 전처리한 데이터를 학습하는 딥러닝 네트워크 구조 부분으로 구성된다. 도 3과 같이 학습 과정에서는 데이터 전처리를 수행한 후, 딥러닝 네트워크 구조를 통해 유전체 데이터를 학습시킨다. 수행과정에서는 학습된 데이터를 기반으로 test sets을 사용하여 암 예측을 수행한다.
- [0025] 데이터 전처리부(110)를 설명하면 다음과 같다.
- [0026] 본 발명에서는 미국 국립보건원에서 진행한 TCGA 프로젝트의 공개된 유전체 데이터들 중에서 BLCA, BRCA, COAD, GBM, KIRC, LGG, LUSC, OV, PRAD, SKCM, THCA, UCEC 등의 12가지 암에 대한 데이터를 사용하여 암을 예측한다. 사용된 총 5,554명의 데이터를 중에서 80%는 training sets이며, 나머지 20%는 validation sets과 test sets이다. 데이터 전처리 과정은 개인별 데이터로 분류하는 데이터 분류과정, 학습 결과에 대한 정확도와 학습 시간을 줄여주는 데이터 선별과정, 학습이 가능한 형태로 변환하는 데이터 변환과정 등으로 구성된다.

Label	Cancer Name	Cancer	Training Sets	Validation Sets	Test Sets	Total
1	방광암	BLCA	314	63	16	393
2	유방암	BRCA	776	157	38	971
3	결장선암	COAD	258	52	13	323
4	교모세포종 다형종	GBM	304	61	16	381
5	피부 흑색종	SKCM	285	58	14	357
6	신장신선명백세포암	KIRC	268	53	14	335
7	뇌하급 교모종	LGG	405	81	21	507
8	폐편평세포암	LUSC	379	76	19	474
9	난소성경화증	OV	344	68	18	430
10	전립선암	PRAD	394	79	20	493
11	갑상선암	THCA	393	79	20	492
12	자궁내막암	UCEC	318	65	15	398
total	12	12	4,438	892	224	5,554

- [0027]
- [0028] 데이터 분류에 있어서, 암과의 연관성 예측을 위한 학습 데이터는 개인별 전체 유전체 데이터를 입력으로 사용한다.
- [0029] 그러나 TCGA 데이터는 개인별로 분류되어 있지 않기 때문에 Sample_Barcode를 활용하여 개인별 전체 유전체 데이터로 분류한다. Sample Barcode는 TCGA 데이터베이스에서 각 생체 샘플 데이터들을 구분하기 위한 고유키이다. 도 4는 TCGA 데이터를 Sample_Barcode을 사용하여 개인별 전체 유전체 데이터로 분류한 것을 나타

낸다.

- [0030] 도 5는 유전체 데이터에서 제외할 변수 정보를 나타낸 도면이다.
- [0031] 데이터 선별은 개인별 전체 유전체 데이터로 분류한 후에, 데이터 선별과정을 수행한다. 데이터 선별과정은 불분명하고 결과에 관계없는 특징들을 제외하여 학습 결과에 대한 정확도와 성능을 높이는 매우 중요한 과정이다.
- [0032] 먼저 TCGA 유전체 데이터에는 다양한 정보가 포함되어 있지만 담배, 몸무게, 술 등의 환경적 요인 및 후천적인 요인은 개인정보 보호에 의해 제공되는 부분이 많지 않으므로 학습데이터에 사용할 변수 정보에서 제외한다.
- [0033] 도 6은 선별된 유전체 데이터의 정보를 나타낸 도면이다.
- [0034] 다음에 선별된 6개의 TCGA 데이터 변수들에서 암 종류를 제외한 5개의 데이터 변수들을 학습데이터 변수로 사용한다. 학습데이터에 사용되는 변수는 염색체 번호, 염색체 위치, 유전자 변이, 변이 종류, 성별 등이다. 학습에 사용된 변수들은 모두 중요한 데이터이지만 그 중 변이 종류는 염기서열이 변하는 경우, 아미노산 서열이 변하는지 변하지 않는지 등에 관한 정보를 담고 있으므로 유전체 돌연변이를 식별하는데 매우 중요한 변수이다.
- [0035] 도 7은 학습데이터에 사용할 변수 정보를 나타내고 있다. 도 7에 도시된 바와 같이, 선별된 6개의 TCGA 데이터 변수들에서 암 종류를 제외한, 5개의 TCGA 데이터 변수들과 12가지 암에 대한 데이터를 학습 데이터로 사용한다.
- [0036] 데이터 변환과정은 데이터 선별과정을 수행한 후에, 데이터들을 딥러닝 네트워크에 입력하기 위하여 학습을 수행할 수 있는 형태로 변환하고 가공하는 과정이다. 우선, 개인 유전체 데이터를 character convolution neural network 기반의 딥러닝 네트워크에 적용하기 위하여 도 8과 같이 개인의 여러 유전체 데이터를 하나의 문장으로 변환한다.
- [0037] 도 8은 개인의 유전체 데이터를 하나의 문장으로 변환하는 과정을 나타낸 도면이다. 도 8에 도시된 바와 같이, 데이터 선별 과정 후, 개인의 유전체 데이터를 하나의 문장으로 변환한다. 이는 Character Convolution Neural Network에 적용하기 위함이다.
- [0038] 개인의 여러 유전체 데이터를 하나의 문장으로 변환한 후에 대문자를 소문자로 바꾸고 one hot 인코딩을 수행한다. 본 발명에서 사용한 데이터는 개인별 전체 유전체 데이터이므로 유전체의 변이는 매우 많이 존재한다. 따라서, 새로운 유전체 변이 정보가 들어와도 데이터를 생성하고 학습을 수행하는데 용이한 character 단위로 데이터를 변환하는 one hot 인코딩을 수행한다.
- [0039] 도 9는 one hot 인코딩을 수행한 예시를 나타낸 도면이다. 도 9처럼 "hi hello"의 문장을 a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, -, ', ;, . . . , .(70 character) 등에 대하여 해당되는 문자를 1, 해당되지 않는 문자를 0으로 one hot 인코딩을 수행한다.
- [0040] 도 10은 개인 유전체 데이터에 대한 데이터 변환을 나타낸 도면이다.
- [0041] 선별된 TCGA 유전체 데이터 변수 정보를 기반으로 one hot 인코딩을 수행을 통해 도 10과 같이 4,438명의 학습 데이터를 구축하게 된다. 구축된 학습데이터는 character convolution neural network 기반의 딥러닝 네트워크에 입력으로 들어가게 된다.
- [0042] 도 11은 12가지 암에 대한 개인 유전체 학습 데이터 구축을 설명하기 위한 도면이다. 도 11에 도시된 바와 같이, 선별된 TCGA 유전체 데이터 변수 정보를 기반으로 데이터 변환을 통해 4,438명의 학습 데이터(Training Sets)를 구축한다.
- [0043] 다음으로 본 실시예에 따른 딥러닝 네트워크 구조를 설명하기로 한다.
- [0044] 도 12는 딥러닝 네트워크 구조 설계를 설명하기 위한 도면이다. 도 12에 도시된 바와 같이, one-hot encoding을 수행한 학습데이터를 character convolution neural network 기반의 딥러닝 구조를 거쳐 유전체 데이터와 암 발생과의 연관성을 예측한 결과를 출력한다.
- [0045] 선별된 TCGA 유전체 데이터 변수 정보를 기반으로 one hot 인코딩을 수행을 통해 구축한 학습데이터가, 도 12의 character convolution neural network 기반의 딥러닝 구조를 거쳐 유전체 데이터와 암 발생과의 연관성을 예측한 결과를 출력한다. 딥러닝 네트워크 모델의 층이 너무 깊은 경우에 gradient vanishing, gradient exploding 등의 문제들로 인해 모델의 성능이 낮아지는 결과를 산출하는 사실이 기존에 발표된 논문들에 의하여 증명되었다. Gradient vanishing은 깊은 모델의 학습을 진행할 때 역전파 과정에서 입력층으로 갈수록 기울기가

점차적으로 작아져 가중치들이 제대로 업데이트되지 않는 경우이다. Gradient exploding은 기울기가 커져 비정상적으로 큰 값으로 발산하는 경우를 말한다. 이러한 문제점을 해결하기 위하여 깊은 층을 학습하기 위한 ResNet 모델이 발표되었다. 본 발명에서는 ResNet 모델의 기본 구조인 shortcut connection을 사용한다. 하나 이상의 레이어를 건너뛰는 shortcut connection을 통해 몇 개의 레이어들을 건너뛰면 gradient vanishing 문제를 해결할 수 있으며, 학습 속도를 개선할 수 있다. 한편, 데이터의 분포를 정규화하는 batch Normalization을 사용함으로써 가중치 초기값의 의존성이 줄어들며, 학습데이터에 대해서는 오차가 감소하지만 실제 데이터에 대해서는 오차가 증가하는 overfitting을 줄이게 된다.

[0046] Convolution neural network는 일반적으로 convolution layer와 pooling layer로 구성되어 있으며, 마지막 단계에 fully connected layer를 적용한다. Convolution layer는 입력 데이터에서 다양한 특징들을 추출하며, pooling layer는 입력 데이터의 크기와 파라미터를 줄여 overfitting을 방지하기 위하여 max pooling과 k-max pooling을 모두 사용하게 된다. Max pooling은 데이터의 일정 구간마다 최댓값을 추출하여 down sampling 한다. K-max pooling은 학습데이터의 각 row에서 가장 값이 높은 k개를 추출하여 down sampling 한다. 본 발명에서 사용하는 데이터 셋은 시퀀싱 데이터이기 때문에 일정 구간에 중요한 데이터 값이 있음에도 불구하고 최댓값이 아니면 손실될 가능성이 있다. 따라서 fully connected layer 전에 k-max pooling을 수행하여 전체 구간에서 특징을 추출하게 된다. Pooling layer를 거친 후, fully connected layer를 사용하여 특징을 분류하게 된다. 마지막으로 활성화 함수인 ReLu를 사용하여 학습시간을 향상시키며, softmax를 수행하여 암 예측 결과에 대한 확률을 추정하게 된다.

[0047] 본 실시예에서 Pooling/2를 하는 이유는 input 크기를 1/2로 줄이는 Down Sampling을 수행하여 특징을 추출한다. 이는 입력 데이터의 크기와 파라미터를 줄이므로 Overfitting을 방지할 수 있다.

[0048] 도 13은 Max Pooling을 설명하기 위한 도면이다. 도 13과 같이, 가장 큰 값을 대표 값으로 추출하는 Max Pooling방법을 사용한다. K-max Pooling을 하는 이유를 살펴보면 다음과 같다. 기존 Max Pooling은 입력 데이터의 일정 구간마다 최대값을 추출하는데, 일정 구간에 중요한 데이터 값이 있음에도 불구하고 최대값이 아니면 손실될 가능성이 있다. K-max pooling의 경우 입력 데이터에서 k개 만큼의 최대값을 추출한다. 사용하는 데이터셋이 시퀀싱 데이터이기 때문에, 전체 구간에서 특징을 추출하는 것이 좋다고 판단한다. 따라서, Fully connected Layer 전에 K-max pooling을 수행하는 이유는, 입력 데이터 전 구간에서 k개 만큼의 최대값을 추출하므로 데이터 손실의 가능성이 적다.

[0049] 도 14는 Max pooling과 K-max Pooling을 비교하기 위한 도면이다.

[0050] 도 15는 Fully connected layer를 설명하기 위한 FC layer이다. 도 15에 도시된 바와 같이, Convolution layer와 pooling 층만으로는 분류를 해결할 수 있는 구조가 아니므로 특징을 분류할 수 있는 FC layer를 사용한다.

[0051] 도 16은 활성화 함수를 나타낸 그래프이다. 본 실시예에 따른 시스템에서 활성화 함수를 ReLu로 사용한 이유는 다음과 같다.

$$f = \begin{cases} (x \leq 0), & f(x) = 0 \\ (x > 0), & f(x) = x \end{cases}$$

[0052] 여기서, $x > 0$ 이면 기울기가 1인 직선이고, $x < 0$ 이면 함수값이 0이 된다.

[0053] 지수함수인 sigmoid에 비하여 학습시간이 빠르므로 사용한다. sigmoid의 Gradient Vanishing 문제를 해결하므로 사용한다. $x < 0$ 인 값들에 대해서는 기울기가 0이지만 유전자 데이터에는 $x < 0$ 이 없으므로 사용한다.

[0054] 마지막 단계에 Softmax를 사용한 이유는 도 17에 도시된 바와 같이, 뉴런의 출력값에 대하여 클래스를 분류하기 위하여 각 클래스에 대한 확률을 추정한다.

$$y_k = \frac{e^{ak}}{\sum_{i=1}^n e^{a_i}}$$

[0055] 여기서, 함수의 합은 1이며, 결과를 확률 값으로 해석한다.

[0056] 도 18은 본 발명의 일 실시예에 따른 backpropagation을 적용한 개요도이다.

[0059] 본 발명에서는 loss가 일정이상 올라가면 backpropagation을 통해 loss를 감소하는 backpropagation을 사용한다. 이를 통해 암 예측 결과에 대한 정확도 향상의 효과를 얻는다. 입력 데이터와 제안한 기법의 딥러닝 네트워크 구조를 적용하여 나온 결과를 validation sets의 결과와 비교한 후에, 도 18과 같이 loss가 0으로 수렴할 때까지 backpropagation 과정을 반복 수행하여 출력 데이터가 원하는 결과가 나오도록 가중치를 수정한다. 학습에 사용된 데이터는 총 4,438명의 유전체 데이터이며, validation sets에 사용된 데이터는 892명의 유전체 데이터이다. Epoch는 150번, batchsize는 8, learning rate 0.0001로 설정하였다.

[0060] 본 실시예에 따른 실험 환경은 다음과 같다.

[0061] <하드웨어> Intel(R) Core(TM) i7-7700 3.60GHz CPU○RAM 16GB○NVIDIA GeForce GTX 1050(V-RAM 2GB)

[0062] <운영체제> Window 10 Pro 64비트

[0063] <개발도구> Pycharm Community Edition 2019.3.4

[0064] <라이브러리> Torch 1.2.0○CUDA 10.0○cuDNN 7.6.5

[0065] 본 발명에서는 BLCA, BRCA, COAD, GBM, KIRC, LGG, LUSC, OV, PRAD, SKCM, THCA, UCEC 등의 12가지 암에 대한 개인별 유전체 데이터를 입력으로 사용하여 전체적인 유전체 패턴을 분석한다. 또한, 점 돌연변이뿐만 아니라 일반적인 여러 쌍의 SNP, DEL, INS 등의 유전자 변이 타입을 적용하며, 성별 정보 등도 포함한다. 12가지 암에 대하여 총 5,554명의 데이터를 사용하였으며, training sets은 4,438명의 데이터를 사용하였다. 학습이 잘 진행되는지 판단하기 위하여 892명의 validation sets을, 학습 결과를 확인하기 위하여 224명의 test sets을 사용하였다. 표 1은 12가지 암에 대하여 사용된 data sets의 명칭과 개수를 나타내고 있다.

[0066] (표1)

Label	Cancer Name	Cancer	Training Sets	Validation Sets	Test Sets	Total
1	Bladder Urothelial Carcinoma	BLCA	314	63	16	393
2	Breast invasive carcinoma	BRCA	776	157	38	971
3	Colon adenocarcinoma	COAD	258	52	13	323
4	Glioblastoma multiforme	GBM	304	61	16	381
5	Skin Cutaneous Melanoma	SKCM	285	58	14	357
6	Kidney renal clear cell carcinoma	KIRC	268	53	14	335
7	Brain Lower Grade Glioma	LGG	405	81	21	507
8	Lung squamous cell carcinoma	LUSC	379	76	19	474
9	Ovarian serous cystadenocarcinoma	OV	344	68	18	430
10	Prostate adenocarcinoma	PRAD	394	79	20	493
11	Thyroid carcinoma	THCA	393	79	20	492
12	Uterine Corpus Endometrial Carcinoma	UCEC	318	65	15	398
Total		12	4,438	892	224	5,554

[0067]

[0068] 암 예측 결과에 대한 평가를 설명하면 다음과 같다.

[0069] 본 발명에서 제안한 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측 결과에 대한 객관적 신뢰성을 평가하기 위하여 기존의 발표된 논문과 같은 TCGA 데이터 셋을 사용하여 정확도, 정밀도, 특이도, 민감도, F-score 등을 비교 평가하였다. 2019년 Yingshuai Sun 외 7명이 Scientific Reports에서 발표한 논문은 정상인 유전자 염기서열을 기반으로 특정 유전자의 염기서열에서 다른 부분을 1로 같은 부분은 0으로 표기하여 데이터를 생성한 후, 딥러닝을 통해 12가지의 암을 식별하였다. 따라서 본 발명에서도 Yingshuai Sun 외 7명이 Scientific Reports에서 발표한 논문과 같은 12가지의 암에 대한 유전체 데이터를 사용하여 암을 예측한다. 예측한 12가지의 암은 BLCA, BRCA, COAD, GBM, KIRC, LGG, LUSC, OV, PRAD, SKCM, THCA, UCEC 등이다. 도 19는 모델의 성능을 평가하는 방법을 나타내고 있다. 도 19에서 TP(True Positive)는 실제 값은 1이며, 모델이 1로 예측한 경우이며, TN(True Negative)는 실제 값은 0이고 모델이 0으로 예측한 경우이다. 또한, FP(False Positive)는 실제 값은 1이지만, 모델이 0으로 예측한 경우이며, FN(False Negative)는 실제 값은 0이지만 모델이 1로 예측한 경우이다.

[0070] 정확도는 대표적으로 사용되는 지표로서 측정하고자 하는 값이 바르게 예측된 정도를 나타내며, 식(1)에 의해 구해진다.

[0071] (식1)

$$Accuracy = \frac{TP + TN}{TN + FN + TP + FP}$$

[0072]

[0073] 민감도는 질병이 있는 사람 중에서 예측 결과가 양성으로 나타나는 경우를 나타내며, 식(2)에 의해 구해진다.

[0074] (식2)

$$Sensitivity = \frac{TP}{FN + TP}$$

[0075]

[0076] 특이도는 질병이 없는 사람 중에서 예측 결과가 음성으로 나타나는 경우를 나타내며, 식(3)에 의해 구해진다.

[0077] (식3)

$$Specificity = \frac{TN}{FP + TN}$$

[0078]

[0079] 정밀도는 예측 결과가 양성으로 판단한 경우 중에서 실제 질병이 있는 사람을 나타내는 경우를 나타내며, 식(4)에 의해 구해진다.

[0080] (식4)

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

[0081]

[0082] F-score는 정밀도와 민감도를 결합한 지표이며, 식(5)에 의해 구해진다.

[0083] (식5)

$$F-score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity}$$

[0084]

[0085] 본 발명의 성능 평가는 표 2와 같이 12가지 암에 대하여 정확도는 74.11%이며, 정밀도는 75.7%이고, 민감도는 73.84%이고, 특이도는 97.61%이며, F-score는 74.1%이다.

[0086] (표 2)

암 이름	암 명칭	Test sets	Precision	Sensitivity	Specificity	F-score
방광암	BLCA	16	58.8	62.5	96.63	60.6
유방암	BRCA	38	70.7	76.3	93.55	73.4
결장선암	COAD	13	100.0	76.9	100.0	87.0
교모세포종 다형종	GBM	16	58.8	62.5	96.63	60.6
피부 흑색종	SKCM	14	68.8	78.6	97.62	73.3
신장신성명백세포암	KIRC	14	68.4	92.9	97.14	78.8
뇌하급 교모종	LGG	21	93.8	71.4	99.51	81.1
폐편평 세포암	LUSC	19	94.4	89.5	99.51	91.9
난소성경화증	OV	18	76.5	72.2	98.06	74.3
전립선암	PRAD	20	64	80.0	95.59	71.1
갑상선암	THCA	20	87.5	70.0	99.02	77.8
자궁내막암	UCEC	15	66.7	53.3	98.09	59.3
Average		224	75.7	73.84	97.61	74.1

[0087]

[0088] 특정 분류 모델의 성능을 평가할 때 사용되는 지표인 confusion matrix는 실제값과 모델이 예측한 값의 관계를

한눈에 알아볼 수 있게 표현한 것이다.

[0089] 도 20은 실제 암에 걸린 수치와 암을 예측한 수치를 나타낸 도면이다. 도 20에서 세로축은 실제 암에 걸린 수치를 나타내며, 가로축은 암을 예측한 수치를 나타낸다. 도 20에서 BRCA의 예측 결과 수치는 UCEC, OV 등이 같이 예측되었기 때문에, BRCA는 UCEC, OV 등과 연관된 암이라고 판단할 수 있다. 다른 암들의 예측 결과 수치도 연관되는 암들과 같이 표현되었기 때문에, 본 발명에서 제안한 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측 모델의 적정성이 입증되었다.

[0090] 또한, 모델의 성능 평가 지표로서 진단의 효율성을 판단하는 Receiver Operating Characteristic(ROC) curve가 있다. ROC curve는 의학 분야에서 진단의 목적으로 많이 사용될 뿐만 아니라 알고리즘의 성능평가 목적으로도 흔히 사용되는 지표이다. ROC 곡선은 기준에 따라 특이도와 민감도의 관계를 표현한 그래프로서 가로축은 1-특이도를 나타내며, 세로축은 민감도를 나타낸다. ROC curve가 좌상단으로 올라갈수록, ROC curve 아래의 면적인 Area Under the Curve (AUC)가 클수록 예측이 잘 되었다고 할 수 있다. 도 21은 본 발명에서 예측한 12가지 암에 대한 ROC curve를 나타내고 있으며, ROC curve 아래의 면적인 AUC가 90% 이상을 나타내면 일반적으로 분류 결과가 우수하다고 인정되고 있다. 따라서 도 21의 ROC curve 아래의 면적인 AUC가 본 발명에서 예측한 12가지 암에 대하여 모두 90% 이상이므로 완벽한 분류 결과를 나타내고 있음을 확인할 수 있다.

[0091] 한편, 본 발명에서 제안한 기법을 Yingshuai Sun 외 7명이 Scientific Reports에서 발표한 논문과 같은 12가지의 암에 대한 유전체 데이터를 사용하여 암을 예측한 결과를 표 3에 나타내었다. 예측한 12가지의 암은 BLCA, BRCA, COAD, GBM, KIRC, LGG, LUSC, OV, PRAD, SKCM, THCA, UCEC 등이다. 표 3과 같이 본 발명에서 제안하는 기법이 2019년 Yingshuai Sun 외 7명이 Scientific Reports에서 발표한 GDL 논문보다 정확도, 민감도, 특이도 등의 결과에서 우수한 결과를 나타내었다. 따라서 본 발명에서는 딥러닝을 활용한 개인의 전체 유전체 데이터와 암 발생과의 연관성 예측에 관한 기법의 효율성이 입증되었다.

[0092] (표 3)

Method	Accuracy	Sensitivity	Specificity
GDL	70.4	65.92	96.27
The Proposed Method	74.11	73.84	97.61

[0093]

[0094] 본 발명의 일 실시예에 따른 시스템의 실험결과에 관하여, 특정 암에 관련된 특정 유전자 데이터를 학습시킨 것이 아니라, 각 암 환자마다 가지고 있는 모든 유전체 패턴을 분석함으로써, 여러 종류의 암에 대한 예측을 할 수 있는 기초 기반이 마련되었다고 판단된다.

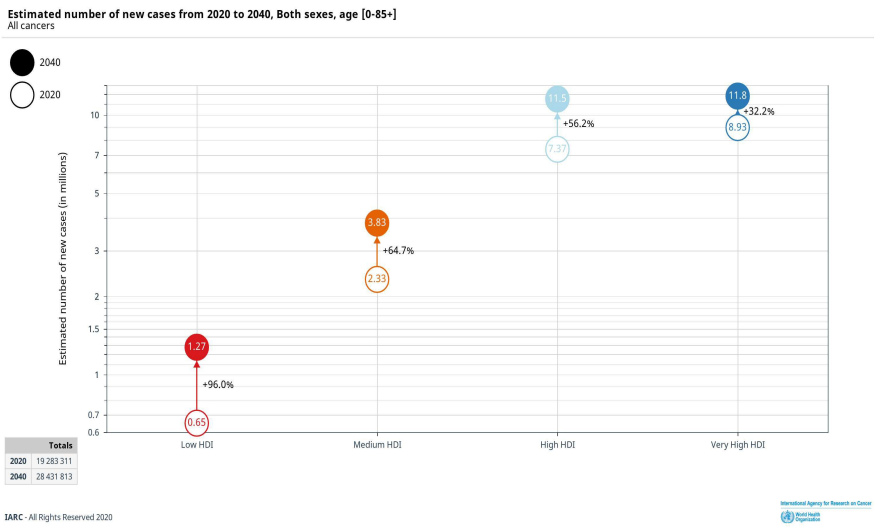
[0095] 자궁암에 Precision과 Recall의 결과가 다른 암보다 낮게 나온다. 이는 BRCA1과 BRCA2 변이 유전자가 유방암과 난소암에 모두 연관이 있으므로 데이터가 많은 유방암에 의해 낮은 결과가 도출되었다고 판단된다.

[0096] 본 발명은 정상인에 대한 참조 서열 정보를 사용하지 않고 개인의 유전체 데이터 정보만을 사용하여 암 질환을 예측한다. 이는 정상인에 대한 참조 서열 정보를 따로 사용하지 않으며, 데이터에 대한 전처리가 비교적 수월하며, 다른 질환에도 적용하기 편리한 효과가 있다.

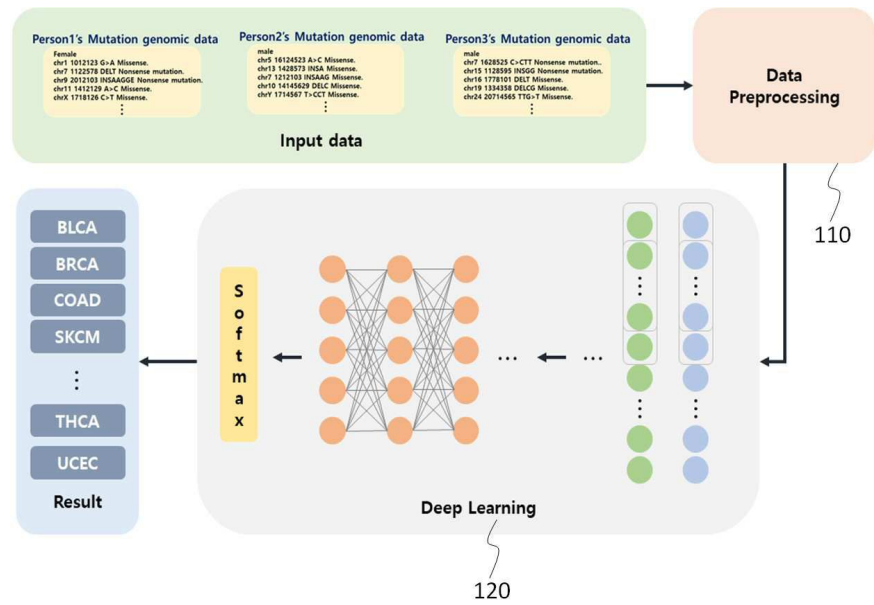
[0097] 이상에서 설명한 본 발명은 전술한 실시예 및 첨부된 도면에 의해 한정되는 것이 아니고, 본 발명의 기술적 사상을 벗어나지 않는 범위 내에서 여러 가지 치환, 변형 및 변경이 가능함은 당업자에게 명백할 것이다.

도면

도면1



도면2



도면6

chrom	chromStart	chromEnd	name	score	strand	thickStart	thickEnd	reserved	blockCount	blockSizes	chromStarts	sampleCount
freq	Hugo_Symbol	Entrez_Gene_Id	Variant_Classification	Variant_Type	Reference_Allele	Tumor_Seq_Allele1	Tumor_Seq_Allele2	dbSNP_RS	dbSNP_Val_Status	days_to_death	cigarettes_per_day	weight
alcohol_history	alcohol_intensity	bmi	years_smoked	height	gender	project_id	ethnicity	Tumor_Sample_Barcode	Matched_Norm_Sample_Barcode	case_id		

<TCGA 유전체 데이터 전체 변수 정보>



chrom	chromStart	name	Variant_Classification	gender	project_id
-------	------------	------	------------------------	--------	------------

<6개의 선별된 TCGA 유전체 데이터 변수 정보>

도면7

염색체 번호	염색체 위치	유전자 변이	변이 종류	성별	암 종류
chrom	chromStart	name	Variant_Classification	gender	project_id
chr19	54906324	G > A	Missense_Mutation	female	TCGA-BRCA
chr5	141303636	C > T	Missense_Mutation	female	TCGA-BRCA
chr11	119156972	G > A	Missense_Mutation	female	TCGA-BRCA
chrX	53549029	C > T	Missense_Mutation	male	TCGA-BRCA
chr7	149104497	C > T	Missense_Mutation	female	TCGA-BRCA
chrX	78755160	insCCTCCTATGAG	Frame_Shift_Ins	female	TCGA-BRCA
chr1	45088038	C > T	Missense_Mutation	female	TCGA-BRCA
chr4	1939673	G > A	Missense_Mutation	male	TCGA-BRCA
chrX	34943721	G > A	Silent	female	TCGA-BRCA
chr4	1939673	G > A	Missense_Mutation	female	TCGA-BRCA
chr14	91812365	A > G	Missense_Mutation	female	TCGA-BRCA
chr12	6814099	C > G	Intron	female	TCGA-BRCA

<학습 데이터에 사용할 변수 정보>

도면8

chrom	chromStart	name	Variant_Classification	gender	project_id	Tumor_Sample_Barcode	Matched_Norm_Sample_Barcode
chr1	20723688	G > A	Missense_Mutation	female	TCGA-BRCA	TCGA-Z7-A896-01A-11D-A41F-09	TCGA-Z7-A896-10A-01D-A41F-09
chr1	22838310	G > A	Missense_Mutation	female	TCGA-BRCA	TCGA-Z7-A896-01A-11D-A41F-09	TCGA-Z7-A896-10A-01D-A41F-09
chr10	12246268	C > T	Nonsense_Mutation	female	TCGA-BRCA	TCGA-Z7-A896-01A-11D-A41F-09	TCGA-Z7-A896-10A-01D-A41F-09
chr19	51498253	G > A	Missense_Mutation	female	TCGA-BRCA	TCGA-Z7-A896-01A-11D-A41F-09	TCGA-Z7-A896-10A-01D-A41F-09
chr19	53149245	G > A	Missense_Mutation	female	TCGA-BRCA	TCGA-Z7-A896-01A-11D-A41F-09	TCGA-Z7-A896-10A-01D-A41F-09
chr2	10123385	G > A	Splice_Site	female	TCGA-BRCA	TCGA-Z7-A896-01A-11D-A41F-09	TCGA-Z7-A896-10A-01D-A41F-09
chr10	133224923	G > T	3'UTR	female	TCGA-BRCA	TCGA-XX-A89A-01A-11D-A36J-09	TCGA-XX-A89A-10A-01D-A36M-09
chr10	1231408	C > T	Silent	female	TCGA-BRCA	TCGA-XX-A89A-01A-11D-A36J-09	TCGA-XX-A89A-10A-01D-A36M-09
chr10	13324923	G > T	3'UTR	female	TCGA-BRCA	TCGA-XX-A89A-01A-11D-A36J-09	TCGA-XX-A89A-10A-01D-A36M-09
chr11	533786	A > C	Missense_Mutation	female	TCGA-BRCA	TCGA-XX-A89A-01A-11D-A36J-09	TCGA-XX-A89A-10A-01D-A36M-09
chr11	16318080	delTTGCT	Intron	female	TCGA-BRCA	TCGA-XX-A89A-01A-11D-A36J-09	TCGA-XX-A89A-10A-01D-A36M-09
chr12	5290506	T > G	3'UTR	female	TCGA-BRCA	TCGA-XX-A89A-01A-11D-A36J-09	TCGA-XX-A89A-10A-01D-A36M-09

→ Person1

→ Person2

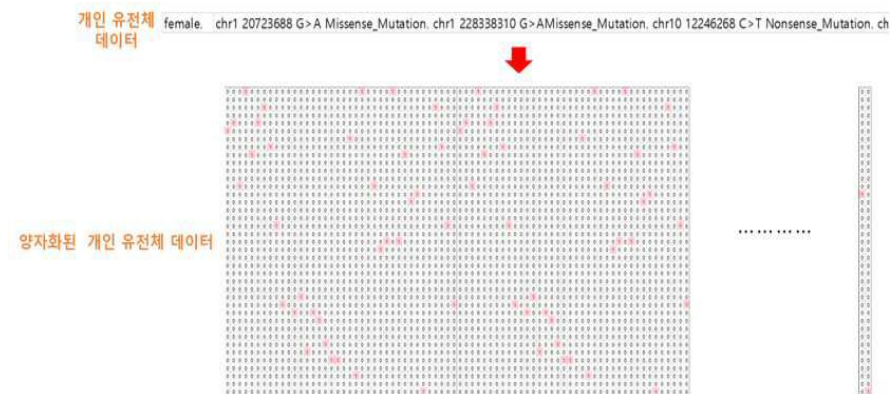


Person1 TCGA-BRCA female. chr1 20723688 G>A Missense_Mutation. chr1 22838310 G>A Missense_Mutation. chr10 12246268 C>T Nonsense_Mutation. chr10 17152036 G>C Intron. chr10 122640347 C>T Silent. chr10 133224923 G>T 3'UTR. chr11 533786 A>C Missense_Mutation.

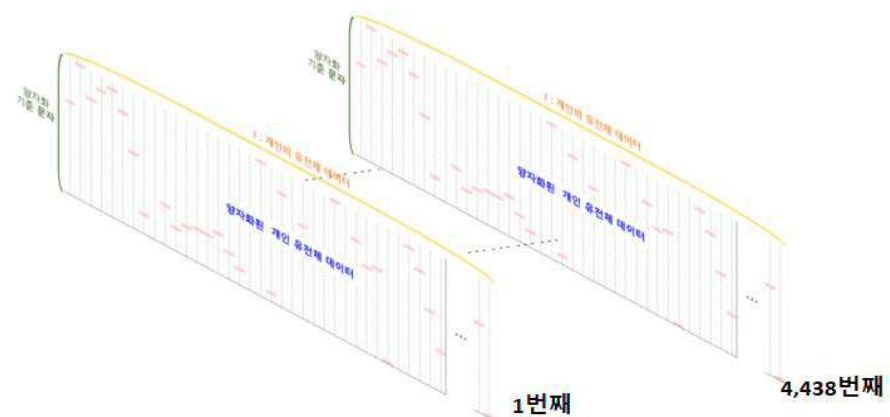
도면9

	hi hello.															
a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
e	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
g	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
h	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
i	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
l	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
m	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
o	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
:	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
.	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

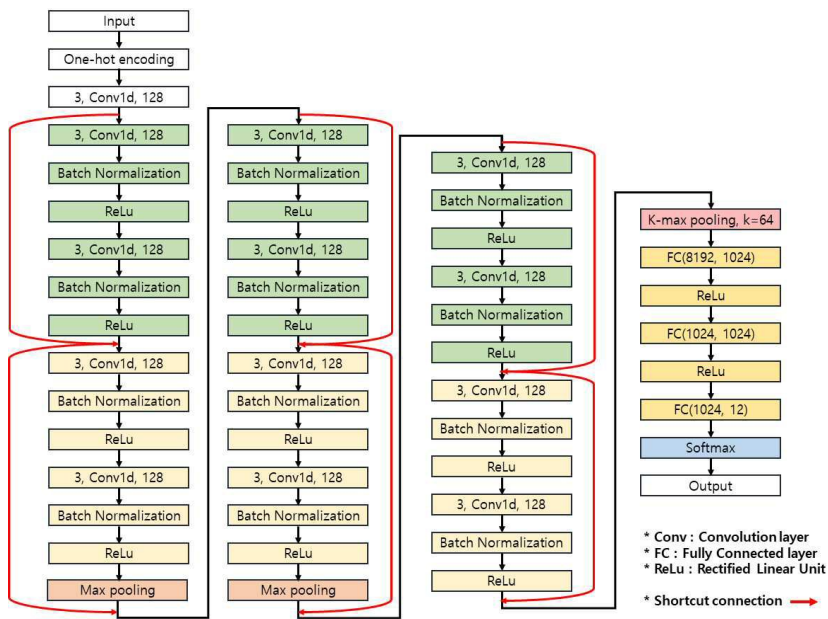
도면10



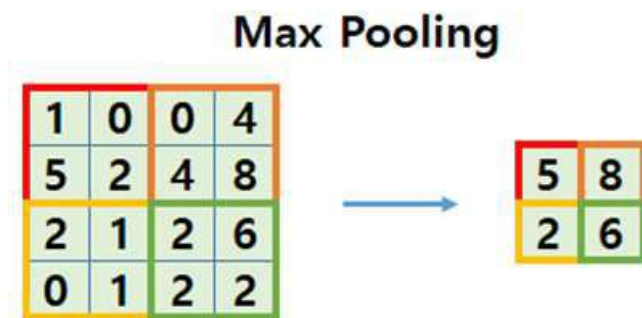
도면11



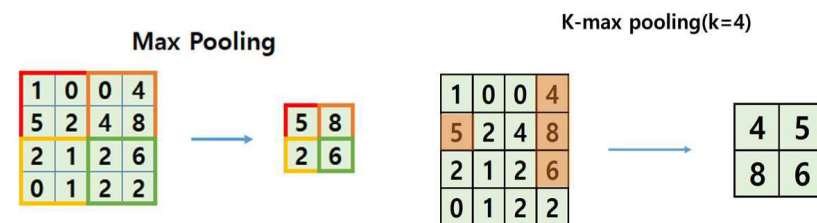
도면12



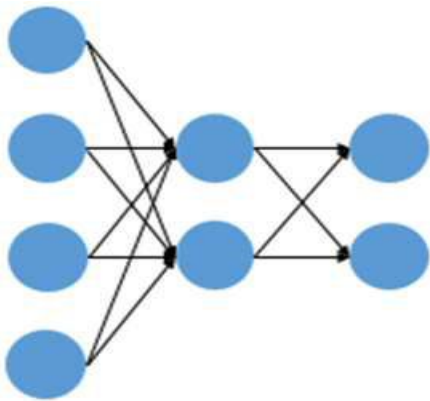
도면13



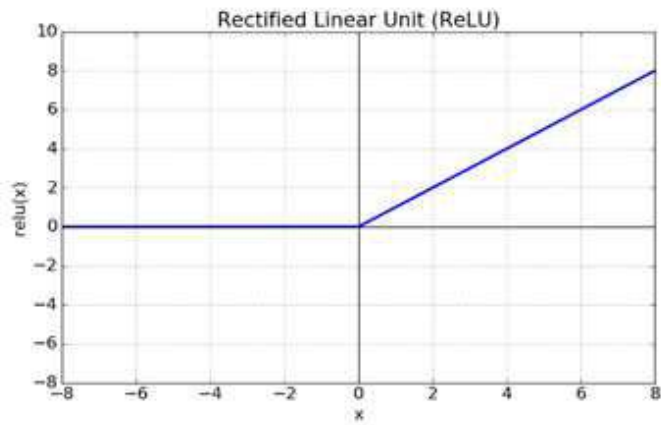
도면14



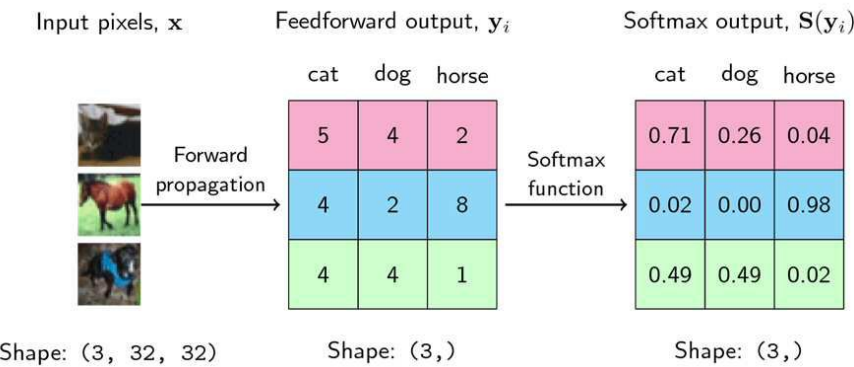
도면15



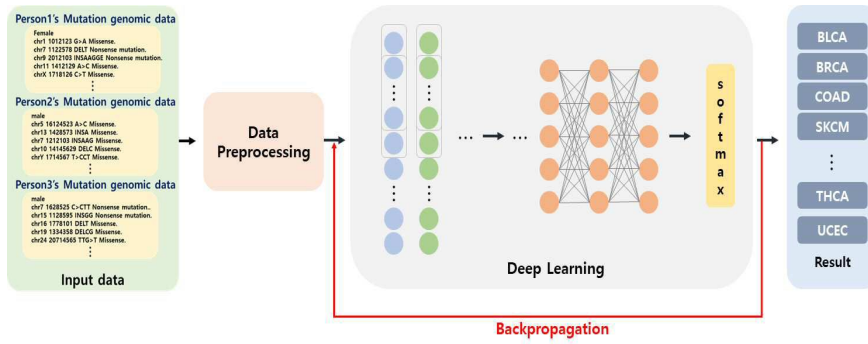
도면16



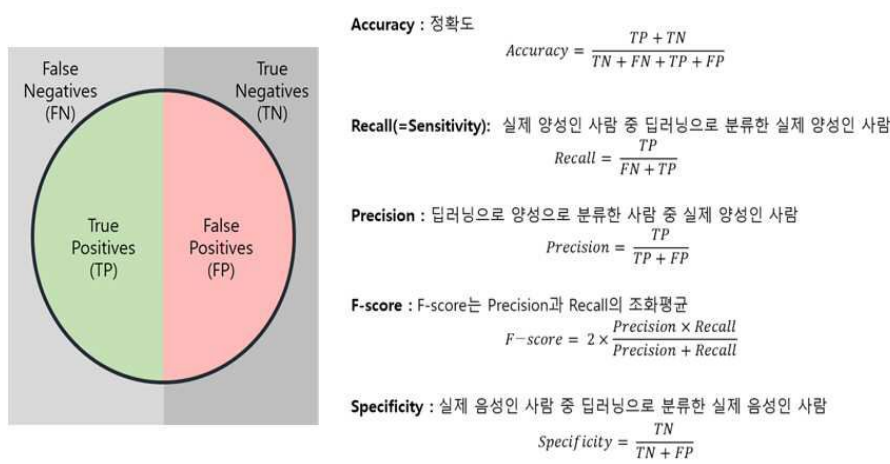
도면17



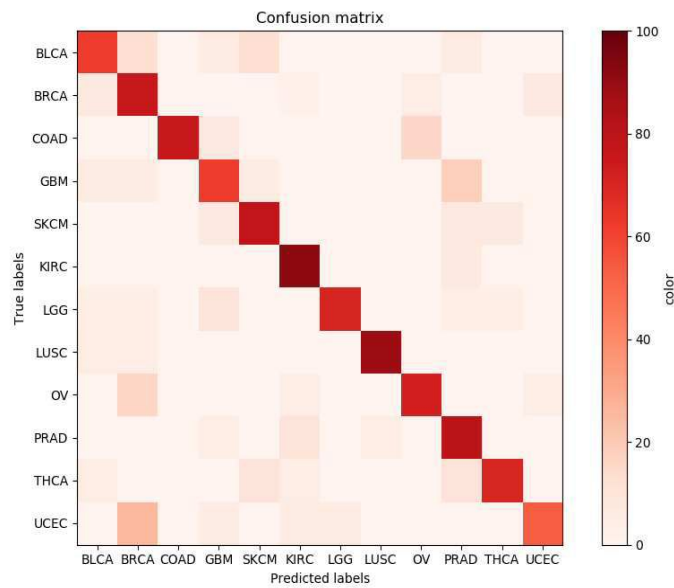
도면18



도면19



도면20



도면21

