

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2018년 8월 9일 (09.08.2018)

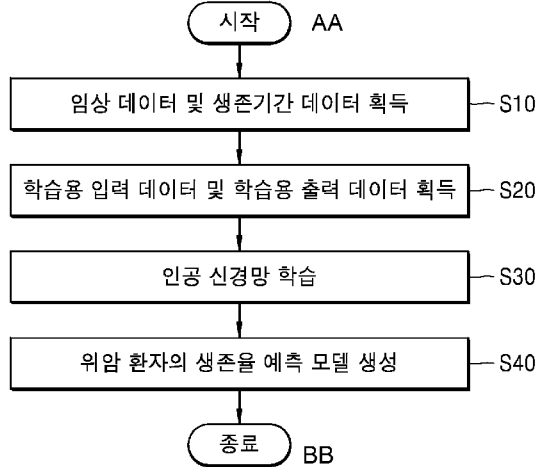


(10) 국제공개번호
WO 2018/143540 A1

- (51) 국제특허분류: *G06F 19/00* (2011.01) *G06N 3/08* (2006.01)
- (21) 국제출원번호: PCT/KR2017/012068
- (22) 국제출원일: 2017년 10월 30일 (30.10.2017)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보: 10-2017-0015128 2017년 2월 2일 (02.02.2017) KR
- (71) 출원인: 사회복지법인 삼성생명공익재단 (SAMSUNG LIFE PUBLIC WELFARE FOUNDATION) [KR/KR]; 04348 서울시 용산구 이태원로55길 48, Seoul (KR).
- (72) 발명자: 서성욱 (SEO, Sung Wook); 06288 서울시 강남구 삼성로 150, 110동 107호, Seoul (KR). 이지연 (LEE, Jeeyun); 06351 서울시 강남구 일원로 81, Seoul (KR).
- (74) 대리인: 리앤목 특허법인 (Y.P.LEE, MOCK & PARTNERS); 06292 서울시 강남구 언주로 30길 13 대림아크로텔 12층, Seoul (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유

(54) Title: METHOD, DEVICE, AND PROGRAM FOR PREDICTING PROGNOSIS OF STOMACH CANCER BY USING ARTIFICIAL NEURAL NETWORK

(54) 발명의 명칭: 인공신경망을 이용한 위암의 예후 예측 방법, 장치 및 프로그램



- S10...Acquire clinical data and data on survival period
- S20...Acquire learning input data and learning output data
- S30...Cause artificial neural network to perform learning
- S40...Generate model for predicting survival rate of stomach cancer patient
- AA...Start
- BB...End

(57) Abstract: A method for predicting a prognosis of stomach cancer by using an artificial neural network according to an embodiment of the present invention comprises the steps of: acquiring data on survival periods after the onset of stomach cancer and clinical data of a plurality of stomach cancer patients; acquiring learning input data and learning output data from the clinical data and the data on survival periods; causing an artificial neural network including an input layer, a hidden layer, and an output layer to perform learning using the learning input data and the learning output data; and generating a model for predicting a survival rate of a stomach cancer patient by using the learned artificial neural network.

WO 2018/143540 A1

럼 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

(57) 요약서: 본 발명의 일 실시예에 따른 인공지능망을 이용한 위암의 예후 예측 방법은, 복수의 위암 환자들의 임상 데이터 및 위암 발병 후 생존 기간 데이터를 획득하는 단계; 상기 임상 데이터와 상기 생존 기간 데이터로부터 학습용 입력 데이터와 학습용 출력 데이터를 획득하는 단계; 상기 학습용 입력 데이터와 상기 학습용 출력 데이터를 이용하여 입력층, 은닉층, 출력층을 포함하는 인공지능망을 학습시키는 단계; 및 상기 학습된 인공지능망을 이용하여 위암 환자의 생존율을 예측하는 모델을 생성하는 단계를 포함한다.

명세서

발명의 명칭: 인공신경망을 이용한 위암의 예후 예측 방법, 장치 및 프로그램

기술분야

[1] 본 발명은 인공신경망을 이용한 위암의 예후 예측 방법, 장치 및 프로그램에 관한 것이다.

[2]

배경기술

[3] 위암은 우리나라 전체 암 중 발생 빈도 1위, 사망 원인 2위일 정도로 많이 발생하는 암이다. 위암은 종양의 주변 구조 침범 정도, 부위 림프절로의 전이, 다른 장기로의 전이 등 진행 정도에 따라 병기가 결정되는데, 이에 따라 치료법과 예후가 달라진다. 병기가 증가할수록 예후는 점점 나빠지고, 특히 말기(Stage 4)에서는 5년 생존율이 3%밖에 되지 않을 정도로 예후가 좋지 않다.

[4] 그러나 같은 병기를 가지더라도 사람에 따라 그 예후는 이질적일 수 있는데, 이는 다양한 환경적, 유전적 원인에서 비롯된다. 따라서 암의 예후를 개인별로 정확하게 예측할 수 있게 해주는 다양한 방법들이 개발되고 있다.

[5] 예컨대 한국등록특허 제10-1415257호에서는 microRNA-196b RNA 및 HOXA10(Homeobox A10) 단백질의 과발현 수준을 측정하여 위암의 예후를 진단하는 방법을 개시하고 있다.

[6] 한편, 한국등록특허 제10-1504818호에서는 여러 유전자 발현 프로파일을 클러스터 분석하여 위암의 예후를 예측하는 시스템을 개시하고 있다.

[7]

발명의 상세한 설명

기술적 과제

[8] 단, 기존의 위암 예후 예측 방법은, 위암 예후를 단순히 저위험군, 중간위험군, 고위험군으로 나누는 것에 불과하여 위암 환자의 생존율 예측을 정확하게 수행할 수 없다는 문제가 있다.

[9] 본 발명은 상기와 같은 문제점을 포함하여 여러 문제점을 해결하기 위한 것으로써, 인공신경망을 이용하여 위암의 예후를 예측하는 방법, 장치 및 프로그램을 제공하는 것을 목적으로 한다.

[10] 그러나 이러한 과제는 예시적인 것으로, 이에 의해 본 발명의 범위가 한정되는 것은 아니다.

[11]

과제 해결 수단

[12] 본 발명의 일 실시예에 따른 인공신경망을 이용한 위암의 예후 예측 방법은, 복수의 위암 환자들의 임상 데이터 및 위암 발병 후 생존 기간 데이터를

- 획득하는 단계; 상기 임상 데이터와 상기 생존 기간 데이터로부터 학습용 입력 데이터와 학습용 출력 데이터를 획득하는 단계; 상기 학습용 입력 데이터와 상기 학습용 출력 데이터를 이용하여 입력층, 은닉층, 출력층을 포함하는 인공신경망을 학습시키는 단계; 및 상기 학습된 인공신경망을 이용하여 위암 환자의 생존율을 예측하는 모델을 생성하는 단계;를 포함한다.
- [13] 일 실시예에 있어서, 상기 학습용 입력 데이터는 상기 복수의 위암 환자들의 분자 유전학적 아형(subtype) 데이터를 포함할 수 있고, 상기 아형은 MSI(microsatellite instable) 아형, MSS/EMT 아형, MSS/TP53+ 아형, MSS/TP53- 아형을 포함한다.
- [14] 일 실시예에 있어서, 상기 입력층은 상기 분자 유전학적 아형 데이터가 입력되는 4개의 노드를 포함할 수 있다.
- [15] 일 실시예에 있어서, 상기 학습용 입력 데이터와 상기 학습용 출력 데이터를 이용하여 상기 인공신경망을 학습시키는 단계는, 상기 학습용 입력 데이터의 각 변수를 2차원 이상의 벡터로 임베딩(embedding)하여 임베딩층을 산출하는 단계를 포함할 수 있다.
- [16] 일 실시예에 있어서, 상기 임상 데이터와 상기 생존 기간 데이터로부터 각각 학습용 입력 데이터와 학습용 출력 데이터를 획득하는 단계는, k-최근접 이웃 알고리즘(k-nearest neighbor algorithm: knn)을 이용하여 결측치(missing data, NaN)를 추가하는 단계를 포함할 수 있다.
- [17] 일 실시예에 있어서, 상기 인공신경망의 상기 은닉층은 적어도 하나의 순환신경망(RNN) 층을 포함할 수 있다.
- [18] 본 발명의 일 실시예에 따른 인공신경망을 이용한 위암의 예후 예측 장치는, 복수의 위암 환자들의 임상 데이터 및 위암 발병 후 생존 기간 데이터를 획득하는 데이터 획득부; 및 상기 임상 데이터와 상기 생존 기간 데이터로부터 학습용 입력 데이터와 학습용 출력 데이터를 획득하고, 상기 학습용 입력 데이터와 상기 학습용 출력 데이터를 이용하여 입력층, 은닉층, 출력층을 포함하는 인공신경망을 학습시키는 인공신경망 학습부; 및 상기 학습된 인공신경망을 이용하여 위암 환자의 생존율을 예측하는 모델을 생성하는 생존율 예측 모델 생성부;를 포함한다.
- [19] 일 실시예에 있어서, 상기 학습용 입력 데이터는 상기 복수의 위암 환자들의 분자 유전학적 아형 데이터를 포함할 수 있고, 상기 아형은 MSI(microsatellite instable) 아형, MSS/EMT 아형, MSS/TP53+ 아형, MSS/TP53- 아형을 포함한다.
- [20] 일 실시예에 있어서, 상기 입력층은 상기 분자 유전학적 아형 데이터가 입력되는 4개의 노드를 포함할 수 있다.
- [21] 일 실시예에 있어서, 상기 인공신경망 학습부는 상기 학습용 입력 데이터의 각 변수를 2차원 이상의 벡터로 임베딩(embedding)하여 임베딩층을 산출할 수 있다.
- [22] 일 실시예에 있어서, 상기 인공신경망 학습부는, k-최근접 이웃 알고리즘(k-nearest neighbor algorithm: knn)을 이용하여 상기 학습용 입력

데이터의 결측치(missing data, NaN)를 추가할 수 있다.

- [23] 일 실시예에 있어서, 상기 인공신경망의 상기 은닉층은 적어도 하나의 순환신경망(RNN) 층을 포함할 수 있다.
- [24] 본 발명의 다른 실시예는 컴퓨터를 이용하여 전술한 인공신경망을 이용한 위암의 예후 예측 방법을 실행하기 위하여 매체에 저장된 컴퓨터 프로그램을 개시한다.
- [25] 전술한 것 외의 다른 측면, 특징, 이점이 이하의 도면, 특허청구범위 및 발명의 상세한 설명으로부터 명확해질 것이다.

[26]

발명의 효과

- [27] 본 발명에 따른 인공신경망을 이용한 위암의 예후 예측 방법, 장치 및 프로그램에 따르면, 위암 환자의 예후를 개개인별로 정확하게 예측할 수 있다. 그뿐만 아니라, 학습된 인공신경망을 이용하여 각 치료 방법에 의한 예후를 시뮬레이션할 수 있으므로 환자별 맞춤형 치료 방법을 결정할 수 있다. 물론 이러한 효과에 의해 본 발명의 범위가 한정되는 것은 아니다.

[28]

도면의 간단한 설명

- [29] 도 1은 본 발명의 일 실시예에 따른 인공신경망을 이용한 위암의 예후 예측 방법을 나타낸 순서도이다.
- [30] 도 2는 ACRG(Asian Cancer Research Group)에서 연구한 위암 환자 집단에서, MSI(microsatellite instable) 아형, MSS/EMT 아형, MSS/TP53+ 아형, MSS/TP53-아형에 따른 생존율의 상관관계를 나타낸 그래프이다.
- [31] 도 3은 본 발명의 일 실시예에 의한 인공신경망의 토폴로지(topology)를 간략하게 나타낸 그림이다.
- [32] 도 4는 본 발명의 일 실시예에 따른 인공신경망의 히트맵(heatmap) 그래프의 일부를 개략적으로 나타낸 그림이다.
- [33] 도 5는 본 발명의 일 실시예에 따른 인공신경망을 이용한 위암의 예후 예측 방법에 따라 위암 환자의 t번째 구간 생존율을 예측하는 모델을 생성하는 방법을 예시한 그림이다.
- [34] 도 6는 본 발명의 일 실시예에 따라 연도별로 순차적으로 학습시킨 인공신경망의 히트맵(heatmap) 그래프의 일부를 개략적으로 나타낸 그림이다.
- [35] 도 7은 트레이닝 중 연도별 생존율 예측 모델의 ROC 그래프의 AUC값을 비교한 그래프이다.
- [36] 도 8은 별도의 테스트 데이터로 연도별 생존율 예측 모델을 검증한 ROC 그래프이다.
- [37] 도 9는 생존율 예측 모델이 예측한 생존율과 실제 생존 비율을 비교한 그래프이다.

- [38] 도 10 내지 도 14는 각각 1년, 2년, 3년, 4년, 5년 후 생존율 예측 모델의 판단 곡선(decision curve)을 나타낸 그래프이다.
- [39] 도 15는 본 발명의 일 실시예에 따른 인공신경망과 비교용 단순 인공신경망의 학습 효과를 나타낸 그래프이다.
- [40] 도 16은 타 지역 데이터를 이용하여 인공신경망을 재학습시키는 방법에 대한 모식도이다.
- [41] 도 17은 본 발명의 일 실시예에 따른 인공신경망을 이용한 위암의 예후 예측 모델인 RSN(Recurrent Survival Network)에 싱가포르 데이터를 추가하여 재학습시킨 모델과 싱가포르 데이터만으로 학습된 모델을 비교한 히트맵 그래프이다.
- [42] 도 18은 원래의 모델, 싱가포르 데이터를 추가하여 재학습시킨 재학습 모델, 싱가포르 데이터만으로 학습시킨 싱가포르 모델의 효과를 비교한 그래프이다.
- [43] 도 19는 본 발명의 일 실시예에 따른 인공신경망을 이용한 위암의 예후 예측 장치의 구성을 개략적으로 나타낸 그림이다.

[44]

발명의 실시를 위한 최선의 형태

- [45] 본 발명은 다양한 변환을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 설명하고자 한다. 본 발명의 효과 및 특징, 그리고 그것들을 달성하는 방법은 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 다양한 형태로 구현될 수 있다.
- [46] 이하의 실시예에서, 제1, 제2 등의 용어는 한정적인 의미가 아니라 하나의 구성 요소를 다른 구성 요소와 구별하는 목적으로 사용된다.
- [47] 이하의 실시예에서, 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다.
- [48] 이하의 실시예에서, 포함하다 또는 가지다 등의 용어는 명세서상에 기재된 특징, 또는 구성요소가 존재함을 의미하는 것이고, 하나 이상의 다른 특징들 또는 구성요소가 부가될 가능성을 미리 배제하는 것은 아니다.
- [49] 이하의 실시예에서, '노드'는 특정 값을 가지고 특정 알고리즘을 거쳐 그 값을 변화시킬 수 있으며, 다른 노드와 연결할 수 있는 추상적 개념의 객체를 의미한다.
- [50] 이하의 실시예에서, 용어 '입력층'은 사용자가 부여한 특정 변수를 가지는 한 개 이상의 노드들의 집합이며, 용어 '출력층'은 사용자가 정한 특정한 절차에 따라서 그 절차의 결과값을 가지는 한 개 이상의 노드들의 집합이고, '은닉층'은 사용자가 정해진 절차를 수행할 때에 임시로 나타나는 중간 결과 및 임시값을 저장하는 한 개 이상의 노드들의 집합을 의미한다.
- [51] 입력층의 노드들과 은닉층의 노드들 사이, 그리고 은닉층의 노드들과 출력층의

노드들 사이에는 각각 링크들이 존재할 수 있으며, 이 링크들은 사용자가 정의한 절차에 의해 부여받는 특정한 가중치를 가질 수 있다.

[52] 이하의 실시예에서 용어 '예후'는 환자의 생존율, 병세의 진행, 회복에 관한 예측을 나타내는 의학용어이다.

[53] 이하, 첨부된 도면을 참조하여 본 발명의 실시예들을 상세히 설명하기로 하며, 도면을 참조하여 설명할 때 같거나 대응하는 구성 요소는 같은 도면부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다.

[54]

[55] 도 1은 본 발명의 일 실시예에 따른 인공신경망을 이용한 위암의 예후 예측 방법을 나타낸 순서도이다.

[56] 본 발명의 일 실시예에 따른 인공신경망을 이용한 위암의 예후 예측 방법은, 복수의 위암 환자들의 임상 데이터 및 위암 발병 후 생존 기간 데이터를 획득하는 단계; 상기 임상 데이터와 상기 생존 기간 데이터로부터 학습용 입력 데이터와 학습용 출력 데이터를 획득하는 단계; 상기 학습용 입력 데이터와 상기 학습용 출력 데이터를 이용하여 입력층, 은닉층, 출력층을 포함하는 인공신경망을 학습시키는 단계; 및 상기 학습된 인공신경망을 이용하여 위암 환자의 생존율을 예측하는 모델을 생성하는 단계;를 포함한다.

[57] 도 1을 참조하면, 복수의 위암 환자들의 임상 데이터 및 위암 발병 후 생존 기간 데이터를 획득하는 단계(S10)가 수행된다.

[58] 본 명세서에서 임상 데이터는, 환자의 나이, 성별 등의 신체적 개인정보와 위암 발생 후의 수술 기록, 재발 여부 등 위암에 관련된 병적 기록 등을 포함한다. 위암 발병 후 생존 기간 데이터는, 이미 사망한 환자의 경우에는 위암 발생을 인지한 시점부터 사망까지의 기간을, 생존하고 있는 환자의 경우에는 위암 발생을 인지한 시점부터 본 발명을 실시하는 시점까지의 기간을 의미할 수 있다.

[59] 임상 데이터 및 위암 발병 후 생존 기간 데이터는 한 개 이상의 병원 또는 지역의 위암 환자들로부터 획득할 수 있다. 임상 데이터는 환자의 의료 영상으로부터 획득되거나, 환자의 검체 검사 결과로부터 획득될 수 있으나 이에 한정되지 않는다. 본 발명자들은 5년 이상 추시 관찰한 삼성서울병원의 1187명의 위암 환자들로부터 임상 데이터 및 생존 기간 데이터를 획득하였다

[60] 이후, 임상 데이터와 생존 기간 데이터로부터 각각 학습용 입력 데이터와 학습용 출력 데이터를 획득하는 단계(S20)가 수행된다.

[61] 학습용 입력 데이터는 후술할 인공신경망을 학습하기 위해 입력층의 노드에 입력될 데이터를 의미한다. [표 1]은 학습용 입력 데이터에 포함될 수 있는 변수 및 이의 분류를 나타낸다.

[62] [표1]

| 변수 | 분류 |
|---|--------------------------------|
| 분자 유전학적 아형 | molecular subtype 1: MSI |
| | molecular subtype 2: MSS/EMT |
| | molecular subtype 3: MSS/TP53+ |
| | molecular subtype 4: MSS/TP53- |
| RAS 시그니처(RAS signature) | 실수값 |
| 성별(sex) | 남자: 1, 여자: 2 |
| 나이 | 실수값 |
| HER2 | 0= negative; 1= positive |
| WHO 암 분류 | 1=w/d adeno |
| | 2=m/d adeno |
| | 3=p/d adeno |
| | 4=signet ring |
| | 5= mucinous |
| | 6=papillary adeno |
| | 7=adenosquamous |
| | 8=undifferentiated ca |
| | 9=hepatoid adenoca |
| | 10=tubular adenoca |
| | 11=others (text) |
| LAUREN 병리적 분류 | 1=intestinal |
| | 2=diffuse |
| | 3=mixed |
| 병리적 소견: 주변 신경 침범 여부 (perineural invasion) | PNI 0=(-), 1=(+) |
| 병리적 소견: 림프관 침범 여부 | inv 0=(-), 1=(+) |

| | |
|--------------------------------|--|
| TNM stage | T |
| | N |
| | M |
| 절제 림프절 개수 | # LN dissected: 정수값 |
| 암 침범 림프절 개수 | # of positive node (+): 정수값 |
| 재수술 위치(revised location) | Cardia |
| | Body |
| | Antrum |
| 수술 방법 TG; total , ST: subtotal | 1=TG, 2=STG |
| 보조적 요법 완수 여부 | 0=completed 1=not completed |
| 보조 요법 종류(ADJ CTx. Description) | CCRT |
| | LF_RT |
| | XP_RT |
| | XP |
| | Others |
| 재발 여부(recurrence) | no=0 yes=1 |
| 재발 위치 | First site of recurrence_liver 0=(-), 1=(+) |
| | First site of recurrence_peritoneal seeding 0=(-), 1=(+) |
| | First site of recurrence_ascites (clinically significant) 0=(-), 1=(+) |
| | First site of recurrence_intraabdominal_LN 0=(-), 1=(+) |
| | First site of recurrence_distant lymph node 0=(-), 1=(+) |
| | First site of recurrence_bone 0=(-), 1=(+) |
| | recurrence sites_others 0=(-), 1=(+) |

[63]

[64] 위의 변수뿐 아니라 K-ras 돌연변이 여부, 수술 날짜, 재발 시기 등과 같은 다양한 임상적 변수 역시 학습용 입력 데이터에 포함될 수 있음은 자명하다.

[65]

이 중, 위암의 분자유전학적 아형은 위암 치료의 마이크로세틀라이트 불안정성(microsatellite instability)을 측정하여 분류한 MSI(microsatellite instable) 아형과 MSS(microsatellite stable) 아형, MSS 아형에 대해 EMT(epithelial-to-mesenchymal transition)을 측정하여 분류한 MSS/EMT 아형과

- MSS/epithelial 아형, MSS/epithelial 아형에 대해 *TP53*(Tumor Protein 53)의 활성을 측정하여 분류한 MSS/*TP53*+ 아형과 MSS/*TP53*- 아형으로 분류된다.
- [66] 본 발명의 일 실시예에 따르면, 학습용 입력 데이터는 복수의 위암 환자들의 분자 유전학적 아형(subtype) 데이터를 포함하고, 상기 아형은 MSI(microsatellite instable) 아형, MSS/EMT 아형, MSS/*TP53*+ 아형, MSS/*TP53*- 아형을 포함한다.
- [67] 상기와 같은 새로운 위암의 분자유전학적 아형을 분류하기 위해, 본 발명자들은 삼성병원(Samsung Medical Center, SMC)에서 위전절제술(total gastrectomy) 또는 위부분절제술(subtotal gastrectomy)을 받은 환자의 일차 종양 시료(n=300)를 입수하였다.
- [68] 먼저, 위암 시료는 마이크로세틀라이트 불안정성(microsatellite instability) 정도에 따라 MSI(microsatellite instable) 아형과 MSS(microsatellite stable) 아형으로 분류될 수 있다.
- [69] MSI 아형의 경우에는 유전자 돌연변이가 많이 발생하며, 상대적으로 암의 진행이 느린 편이다. 또한 MSI 아형 환자의 약 60%가 암 1기 또는 2기에 해당하였으며, 평균 생존기간은 100.9 개월로 가장 길게 나타나며 로렌(Lauren) 분류에서 장형(intestinal)이 많은 비율로 속해 있다.
- [70] MSS 아형은, EMT(epithelial-to-mesenchymal transition)을 측정하여 MSS/EMT 아형과 MSS/epithelial 아형으로 분류할 수 있다. EMT 측정으로 간엽(mesenchymal)과 유사한 것으로 측정되는 경우 이를 MSS/EMT 아형으로 분류하며, 상피(epithelial)와 유사한 것으로 측정되는 경우 이를 MSS/epithelial 아형으로 분류한다.
- [71] MSS/EMT 아형은 로렌 분류에서 확산성(diffuse) 종양을 주로 포함하고, 유전자 돌연변이가 거의 나타나지 않으며, 위암 환자의 예후가 좋지 않다. 또한 이 아형은 낮은 연령에서 발견되며 암의 진행이 빠르고 가장 높은 재발률(63%)을 나타낸다.
- [72] MSS/epithelial 아형의 경우에는 *TP53*(Tumor Protein 53)의 활성을 측정하여 MSS/*TP53*+ 아형과 MSS/*TP53*- 아형으로 분류할 수 있다. *TP53*의 활성도가 높은 경우 MSS/*TP53*+ 아형으로, 활성도가 낮은 경우 MSS/*TP53*- 아형으로 분류된다. MSS/*TP53*+ 아형 및 MSS/*TP53*- 아형은 4종류의 아형 중 중간 정도의 예후와 재발률을 나타낸다. 구체적으로, MSS/*TP53*+ 아형의 경우에는 장형 위암이 많다. MSS/*TP53*- 아형은 분석 대상자 중 가장 많은 환자가 갖는 아형이며, *TP53*의 기능이 소실되어 MSS/*TP53*+ 아형보다 위암 환자의 예후가 좋지 않다.
- [73]
- [74] 도 2는 ACRG(Asian Cancer Research Group)에서 연구한 위암 환자 집단에서, MSI(microsatellite instable) 아형, MSS/EMT 아형, MSS/*TP53*+ 아형, MSS/*TP53*- 아형에 따른 생존율의 상관관계를 나타낸 그래프이다.
- [75] 도 2에서 보듯, MSI 아형이 가장 좋은 생존율을 보였고, 그 다음이 MSS/*TP53*+ 및 MSS/*TP53*- 이며 MSS/EMT 아형이 가장 좋지 않은 생존율을 보였다(log-rank,

P = 0.0004). 즉 위암 환자의 생존율은 MSI 아형이 가장 높고 MSS/EMT 아형이 가장 낮은 것으로 확인되므로, 유전자 분석을 통한 아형 분석에 따라 생존율의 예측이 가능하다. 따라서, 위와 같은 위암의 분자유전학적 아형을 입력값으로 하여 인공신경망을 학습시키는 경우, 인공신경망을 이용한 위암의 예후 예측 정확도가 높아지게 된다.

[76] [표 1]에 예시된 학습용 입력 데이터 중에서, WHO 종양 종류, 분자 유전학적 아형 등과 같이 3개 이상으로 카테고리화 할 수 있는 변수는 one-hot encoding 기법을 이용하여 벡터로 변환할 수 있다. 성별, 암 재발 여부와 같이 2개의 분류밖에 없는 경우에는, 두 개의 분류를 0, 1 또는 1, 2로 라벨링하여 하나의 값으로 변환시킬 수 있다. RAS 시그니처 등과 같은 정량적 변수는 정규화(normalization)한 후 가공하여 하나의 수로 변경시킬 수 있다. 이를 통해 학습용 입력 데이터를 수학적으로 처리할 수 있다.

[77] 예컨대, 아래의 [표 2]와 같은 임상 데이터를 가지는 환자 A가 있다고 가정하자. [표 2]에서는 [표 1]에 나타난 변수 중 일부만을 기재하였다.

[78] [표2]

| 변수 | 분류 | 값 |
|-------------------------|--------------------------------|-----|
| 분자 유전학적 아형 | molecular subtype 1: MSI | 1 |
| | molecular subtype 2: MSS/EMT | 0 |
| | molecular subtype 3: MSS/TP53+ | 0 |
| | molecular subtype 4: MSS/TP53- | 0 |
| RAS 시그니처(RAS signature) | 실수값 | 0.9 |
| 성별(sex) | 남자: 1, 여자: 2 | 1 |
| 나이 | 실수값 | 42 |
| HER2 | 0= negative; 1= positive | 1 |

[79]

[80] 이때 학습용 입력 데이터가 입력되는 인공신경망의 노드에 환자 A의 분자 유전학적 아형 정보를 입력할 때, 예컨대 [MSI 노드, MSS/EMT 노드, MSS/TP53+ 노드, MSS/TP53- 노드] = [1, 0, 0, 0] 과 같은 형태로 입력한다. 즉 인공신경망의 입력층은 상술한 분자 유전학적 아형 데이터를 입력하기 위한 4개의 노드를 포함할 수 있다.

[81] 이와 달리, MSI = 1, MSS/EMT = 2, MSS/TP53+ = 3, MSS/TP53- = 4와 같이 라벨링하여 입력층의 [아형 노드] = [1]과 같이 입력하게 되면, 은닉층에서 각 분류에 대한 정보의 손실이 일어나게 된다.

[82] 한편, 성별과 같이 두 개의 분류밖에 없는 경우에는, 성별 정보를 학습용 입력 데이터에 입력할 때 [성별 노드] = 0 또는 1과 같은 형태로 입력한다. 즉 두 개의 분류 밖에 없는 변수의 경우 하나의 노드를 할당한다. 이와 같은 경우에는 노드에 연결된 링크에서의 웨이트(weight)를 통해 성별이 환자의 생존율에

- 미치는 영향을 파악할 수 있으므로, 2개의 노드를 할당하지 않아도 무방하다.
- [83] 한편, 본 발명의 일 실시예에 따르면, 학습용 입력 데이터는 위암 환자의 생존율 또는 사망률 데이터를 포함할 수 있다. 예컨대 각각의 위암 환자의 임상 데이터로부터 얻은 i 개의 데이터가 있는 경우, 학습용 입력 데이터는 상기 위암 환자의 전년도 생존율 데이터 및 전년도 사망률 데이터를 포함한 $i+2$ 개의 데이터 세트일 수 있다. 이에 대하여는 후술한다.
- [84] 학습용 출력 데이터는 인공신경망을 학습하기 위해 출력층의 노드에 입력될 데이터를 의미한다. 이러한 학습용 출력 데이터는 환자의 위암 발병 후 어느 기간 이내의 생존 여부 정보를 포함할 수 있다.
- [85] 학습용 입력 데이터와 학습용 출력 데이터를 획득하는 단계 후에는, 학습용 입력 데이터와 학습용 출력 데이터를 이용하여 인공신경망을 학습시키는 단계(도 1, S30)가 수행된다.
- [86]
- [87] 도 3은 본 발명의 일 실시예에 의한 인공신경망의 토폴로지(topology)를 간략하게 나타낸 그림이다. 인공신경망은 여러 개의 노드를 가지는 입력층, 1개 이상의 은닉층과 출력층을 가진다. 본 발명의 일 실시예에 따른 인공신경망(Recurrent Survival Network, 이하 RSN으로도 명명한다)은 8개의 은닉층을 가지나 본 발명이 이에 제한되는 것은 아니다.
- [88] 인공신경망의 입력층은 n_{in} 개의 노드를 가진다. 도 3에서는 입력층의 노드가 49개인 것을 예시하였으나 이에 제한되는 것은 아니다. 각각의 노드에는 위암 환자 각각의 임상적 변수에 따른 분류를 수학적으로 처리한 값이 입력된다. 이때 입력층은 마치 $n_{in} \times 1$ 행렬과 같은 형태를 가지게 된다.
- [89] 일 실시예에 따르면, 학습용 입력 데이터와 학습용 출력 데이터를 이용하여 인공신경망을 학습시키는 단계는, 학습용 입력 데이터의 각 변수를 2차원 이상의 벡터로 임베딩(embedding)하여 임베딩층을 산출하는 단계를 포함할 수 있다.
- [90] 입력층의 각 노드에 입력되는 학습용 입력 데이터의 각 변수는 1차원 실수값을 가지는데, 이를 임베딩(embedding)하여 2차원 이상의 벡터로 만들 수 있다. 예를 들어 위의 [표 2]의 환자 A의 MSI 노드의 값인 1을 공지된 임베딩 방법을 이용하여 아래와 같이 9차원 벡터로 변환할 수 있다.
- [91] [MSI 노드 임베딩 벡터] = [0.1 0.3 -0.1 -0.5 -0.7 0.2 0.3 0.3 -0.1]
- [92] 이와 비슷하게 각각의 노드의 값을 9차원으로 임베딩하면, 예컨대 아래와 같은 행렬을 얻을 수 있다.

[93]

$$[\text{환자 A 임베딩 행렬}] = \begin{bmatrix} 0.1 & 0.3 & -0.1 & -0.5 & -0.7 & 0.2 & 0.3 & 0.3 & -0.1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0.2 & -0.9 & -0.1 & -0.3 & 0.1 & 0.2 & 0.5 & -0.2 \\ 0.2 & 0.3 & 0.6 & -0.8 & -0.1 & 0.1 & 0.3 & 0.4 & -0.9 \\ 0.1 & 0.4 & -0.7 & -0.9 & -0.3 & 0.4 & 0.7 & 0.2 & -0.5 \\ 0.5 & 0.5 & -0.7 & -0.1 & -0.9 & 0.9 & 0.8 & 0.1 & -0.4 \\ 0.9 & 0.8 & -0.3 & -0.4 & -0.2 & 0.9 & 0.6 & 0.5 & -0.4 \end{bmatrix}$$

[94]

[95] 상기 벡터 및 행렬의 차원 및 각 성분의 값은 본 발명의 이해를 돕기 위하여 예시적으로 정한 것일 뿐 본 발명의 권리범위를 제한하는 것은 아니다.

[96] 본 발명의 실시예에서는 49개의 노드마다 임베딩을 통해 각 실변수(real variable)를 32차원의 벡터로 치환시켰고, 이에 따라 49x32의 행렬 형태의 임베딩층이 산출되었다. 원래 임베딩은 자연어 처리(Natural Language Processing) 분야에서 단어(word)를 수치화하기 위해 고안된 개념이나, 본 발명에서는 각 실변수(real variable)의 수치 그 자체 역시 임베딩을 통해 벡터화하였다. 각 환자의 학습용 입력 데이터를 임베딩하게 되면, 벡터 연산을 통해 각 환자 사이의 유사도를 측정할 수가 있고, 각 변수에 내재한 정보의 손실 없이 데이터를 통합적으로 처리할 수 있게 된다.

[97] 본 발명의 일 실시예에 따르면, 임상 데이터와 생존 기간 데이터로부터 각각 학습용 입력 데이터와 학습용 출력 데이터를 획득하는 단계는, k-최근접 이웃 알고리즘(k-nearest neighbor algorithm: knn)을 이용하여 결측치(missing data, NaN)를 추가하는 단계를 포함할 수 있다.

[98] 예컨대, 아래의 [표 3]과 같은 임상 데이터를 가지는 환자 A, B, C가 있다고 가정하자. 환자 C의 경우에는 HER2 검사를 받지 않아, HER2에 해당하는 값이 결측된(missing, NaN) 상태이다.

[99] [표3]

| 분류 | 환자 A | 환자 B | 환자 C |
|---------------------------------|------|------|------|
| molecular subtype 1: MSI | 1 | 0 | 0 |
| molecular subtype 2: MSS/EMT | 0 | 1 | 1 |
| molecular subtype 3: MSS/TP53+ | 0 | 0 | 0 |
| molecular subtype 4: MSS/TP53- | 0 | 0 | 0 |
| RAS 시그니처 | 0.9 | 0.2 | 0.1 |
| 성별 (남자: 1, 여자: 2) | 1 | 2 | 2 |
| 나이 | 42 | 68 | 65 |
| HER2 (0= negative; 1= positive) | 1 | 0 | NaN |

[100] 이때 환자 C의 임상 데이터가 환자 A와 환자 B 중 누구에게 더 가까운지는

예컨대 각 환자의 학습용 입력 데이터 벡터의 거리를 통해 판별할 수 있다. 예시한 [표 3]의 경우에는 환자 C의 임상 데이터가 환자 A보다는 환자 B에 가까우므로, 환자 C의 HER2 값에 1을 부여할 수 있다.

- [101] 실제로는 비교해야 하는 환자의 수가 많으므로, 위의 예시는 knn 알고리즘을 설명하기 위해 상황을 단순화한 것에 불과할 뿐 반드시 실제의 상황을 반영하는 것은 아니다. 이때 공지된 다양한 knn 알고리즘이 있으므로 본 명세서에서는 자세한 기재를 생략한다.
- [102] 예컨대 본 발명의 인공신경망의 입력층에 입력되는 학습용 입력 데이터의 항목이 다른 지역 또는 병원의 임상 데이터에는 없는 경우, knn 알고리즘을 이용하여 결측된 항목을 추가할 수 있다. 따라서 결측된 데이터가 있는 타 지역 데이터를 추가시켜 인공신경망을 재학습시키는 것이 가능한데, 이에 대하여는 후술한다.
- [103] 한편, 인공신경망의 출력층은 n_{out} 개의 노드를 가진다. 도 3에서는 출력층의 노드가 2개인 것을 예시하였으나 이에 제한되는 것은 아니다.
- [104] 출력층은 환자의 N+1년차 (또는 반기, 분기, 월, 일 등 임의의 시간 단위일 수 있으며, N은 0 이상의 정수) 생존율 및 사망률을 나타내는 노드를 포함할 수 있다. 예컨대 인공신경망이 위암 환자의 위암 발병 후 2년 후의 생존율을 예측하도록 학습되는 경우, 출력층의 노드에는 특정 환자의 위암 발병 후 2년 후의 생존 여부를 나타내는 값이 입력될 수 있다. 예컨대, 특정 환자가 위암 발병 후 2년 전에 사망한 경우, [생존율 노드, 사망률 노드] = [0, 1] 이 되고, 위암 발병 후 2년 후에 사망한 경우에는 [생존율 노드, 사망률 노드] = [1, 0] 이 된다.
- [105] 다만 본 발명의 일 실시예에 따르면, 환자가 사망한 경우의 [생존율 노드, 사망률 노드]를 [0, 1]로 처리하지 않고, 랭킹화하여 스코어를 부여할 수 있는 처리 방법이 제안된다. 본 발명의 일 실시예에서 사망한 환자의 [생존율 노드, 사망률 노드] = [p, 1-p]일 수 있고, 여기서 p에는 0이 아닌 스코어 값이 부여될 수 있다. 이때 스코어는 사망한 환자의 생존 기간에 비례하도록 부여될 수 있다. 여기서 생존 기간은, 적어도 월 단위로 구분될 수 있다. 예컨대, N+1년차에 3개월 생존한 환자의 경우 N+1년차 생존율 스코어는 3/12이고, [생존율 노드, 사망률 노드]의 값은 [3/12, 1-3/12]=[0.25, 0.75]일 수 있다.
- [106] 한편 이때, 입력층에는 [N년차 생존율 노드, N년차 사망률 노드]가 포함될 수 있다. 즉 본 발명에 따르면 위암 환자의 임상 데이터 정보 및 N년차 생존율 및 N년차 사망률 정보를 포함하는 학습용 입력 데이터를 입력층에 입력하고, 환자의 생존 기간에 비례하는 N+1년차 생존율 및 사망률 정보를 포함하는 학습용 출력 데이터를 출력층에 입력하여 인공신경망을 학습시키게 된다.
- [107] 인공신경망의 학습시에는 은닉층(hidden layers)이 활용된다. 각각의 은닉층의 노드는 인접한 다른 은닉층의 노드와 서로 완전히 연결(fully connected) 될 수 있다. 본 발명의 일 실시예에서는, 한 층의 LSTM(Long Short Term Memory) 알고리즘을 이용한 RNN(Recurrent Neural Network)층을 포함하는 8개의

은닉층을 사용하여 인공신경망을 학습시키거나 은닉층의 개수 및 알고리즘의 종류 등은 이에 제한되지 않는다.

- [108] 다시 도 1을 참조하면, 인공신경망 학습시키는 단계(S30) 후에는, 학습된 인공신경망을 이용하여 위암 환자의 생존율을 예측하는 모델을 생성하는 단계(S40)가 수행된다. 인공신경망이 학습된 후에는 각 노드에 대응하는 계수(weight)가 생존율 예측에 최적화되었으므로, 임의의 위암 환자의 임상 데이터로부터 얻은 입력 데이터를 인공신경망의 입력층에 입력하여 나온 출력값을 통해 환자의 생존율을 예측할 수 있다.

[109]

발명의 실시를 위한 형태

[110] <실시예 1>

- [111] 본 연구진은 삼성서울병원의 1187명의 환자에서 얻은 19개 이상의 임상 변수와 5년 이상의 추시 관찰 결과에 대한 데이터를 토대로 인공신경망을 구축하였다. 학습용 데이터는 총 데이터 중 85% (1009명)를, 테스트용 데이터는 나머지 15% (178명)를 사용하였고, 이는 100번에 걸쳐 임의로 재추출(resampling)되었다. 학습용 데이터 중 15%는 교차 검증(cross-validation)을 위해 다시 분류되었다.
- [112] 최적화된 모델은 각각의 학습 데이터마다 30번 학습되었고, 테스트 데이터를 통해 성능을 평가하였다.

[113]

- [114] 도 4는 본 발명의 일 실시예에 따른 인공신경망의 히트맵(heatmap) 그래프의 일부를 개략적으로 나타낸 그림이다.

[115] 그래프 100은 입력층에 학습용 입력 데이터가 입력된 상태를 나타낸다. 그래프 100의 히트맵(heatmap)의 가로축은 각 위암 환자의 일련번호이고, 세로축은 인공신경망 입력층의 각 노드에 해당한다. 일 실시예에서 입력층의 노드는 총 49개로, [표 1]에 나타난 임상 데이터로부터 얻은 47개의 노드 및 2개의 생존율, 사망률 노드를 포함한다. 각 노드에 해당하는 값은 색상의 농도로 표시된다.

- [116] 이후, 각 위암 환자의 데이터는 임베딩 과정을 통해 행렬로 표현된다. 일 실시예에서 위암 환자의 데이터는 49×32 형태의 행렬로 각각 표현된다. (그래프 101, 102, 103...) 이후, 각 노드에 대한 계수의 학습이 이루어진다. 학습의 결과는 생존 또는 사망으로 라벨링되며, 최종적으로는 소프트맥스(softmax) 함수를 통해 생존 확률로 표현된다. 이때 그래프 130을 참조하면 복수의 위암 환자들에 대한 1년차 생존율이 2개 노드로 표시되었다. 즉, 1년차 생존율 예측 인공신경망은 총 49개의 노드값(그래프 110에 도시)을 총 2개의 노드값(그래프 130에 도시)으로 수렴시킨다.

- [117] 본 발명의 일 실시예에 따른 인공신경망의 은닉층은 적어도 하나의 순환신경망(RNN) 층을 포함할 수 있고, RNN층은 LSTM(Long short term memory) 알고리즘을 이용할 수 있다.

[118]

[119] 도 5는 본 발명의 일 실시예에 따른 인공신경망을 이용한 위암의 예후 예측 방법에 따라 위암 환자의 t 번째 구간 생존율을 예측하는 모델을 생성하는 방법을 예시한 그림이다.

[120] 본 발명의 일 실시예에 따르면, 생존율을 예측하는 모델을 생성하는 단계는 시간 구간(time interval) 별로 상기 인공신경망을 학습시키는 단계를 포함할 수 있다. 상기 시간 구간은 연, 반기, 분기, 월 등 다양할 수 있으나 이하에서는 연(year)을 예시로 설명한다. 예를 들어, 인공신경망은 위암 환자의 임상 데이터로부터 위암 환자의 연차 별 생존율, 예컨대 발병 후 1년부터 5년 후까지의 생존율을 예측하도록 학습될 수 있다.

[121] 본 발명의 일 실시예에 따르면, 생존율을 예측하는 모델을 생성하는 단계는, 상기 임상 데이터 및 상기 복수의 위암 환자들의 t 번째 구간 생존 기간 데이터를 이용하여 t 번째 구간 생존율 예측 모델(SM_t)을 생성하는 단계; 및 상기 t 번째 구간 생존율 예측 모델로부터 얻은 t 번째 구간 생존율 예측 데이터(S_t) 및 상기 복수의 위암 환자들의 $t+1$ 번째 구간 생존 기간 데이터를 이용하여 $t+1$ 번째 구간 생존율 예측 모델(SM_{t+1})을 생성하는 단계를 포함할 수 있다.

[122] 본 발명의 일 실시예에 따르면 위암 환자의 1, 2, ..., t , $t+1$ 번째 구간 별 생존율이 예측되는데(t : 자연수), 이때 $t+1$ 번째 구간에서의 생존율 예측을 위해 t 번째 구간에서의 생존율 예측 결과 데이터가 이용된다. 즉, 각 구간 별 생존율 예측이 귀납적, 순차적인 방식으로 이루어지게 된다.

[123] 도 5를 참조하면, 1년 후 생존율 예측 모델(SM_1)과, t 년 후 생존율 예측 모델(SM_t)이 도시되었다. 이때 초기 임상 데이터(X_1) 및 생존율 초기값(S_0)을 입력하였을 때 1년 후 생존율(S_1)을 출력할 수 있는 입출력함수인 1년 후 생존율 예측 모델(SM_1)이 인공신경망을 학습시켜 생성된다.

[124] 이때 인공신경망의 입력층에 입력되는 학습용 입력 데이터는 초기 임상 데이터(X_1) 및 생존율 초기값(S_0)을 포함한다. 1년 후 생존율 예측 모델의 입력이 되는 초기 임상 데이터(X_1)는 초진 시의 임상 데이터일 수 있다. 생존율 초기값(S_0)은 예컨대 1로 설정될 수 있다.

[125] 1년 후 생존율 예측을 위한 학습용 출력 데이터에는, 환자의 생존 기간 데이터로부터 얻은 1년 후 생존 여부 데이터가 이용된다. 예컨대 어떤 환자 D 가 위암 발병 15개월 후 사망한 경우, 발병 후 1년 후 시점에는 생존하였으므로 출력층의 [생존율 노드, 사망률 노드]에 출력될 값과 비교될 학습용 출력 데이터는 [1, 0]이 된다. 인공신경망은 이러한 학습용 입력 데이터 및 학습용 출력 데이터를 이용하여, 위암 환자의 1년 후 생존율을 예측할 수 있도록 학습된다.

[126] 다음으로, 2년 후 임상 데이터(X_2) 및 1년 후 생존율 예측 결과값(S_1)을 입력하였을 때 2년 후 생존율(S_2)을 출력할 수 있는 입출력함수인 2년 후 생존율 예측 모델(SM_2)이 인공신경망을 학습시켜 생성된다. 이때 인공신경망의 입력층에 입력되는 학습용 입력 데이터는 2년 후 임상 데이터(X_2) 및 1년 후

생존율 예측 결과값(S_t)을 포함한다.

[127] 학습용 출력 데이터에는, 환자의 생존 기간 데이터로부터 얻은 2년 후 생존 여부 데이터가 이용된다. 예컨대 어떤 환자 D가 위암 발병 후 15개월 후 생존한 경우, 발병 후 2년 후 시점에는 사망하였으므로 출력층의 [생존율 노드, 사망률 노드]에 출력될 값과 비교될 학습용 출력 데이터는 [0, 1]이 될 수 있다.

[128] 다만 본 발명의 일 실시예에 따르면, 환자가 사망한 경우 학습용 출력 데이터를 상기와 같이 [0, 1]로 처리하지 않고, 랭킹화하여 스코어를 부여할 수 있는 처리 방법이 제안된다.

[129] 일 실시예에 따르면, t번째 구간 생존율 예측 모델(SM_t)을 생성하는 단계는, 상기 t번째 구간 생존 기간 데이터에 대하여 생존 기간에 따른 스코어를 부여하는 단계를 더 포함할 수 있다. 즉 본 실시예에서 학습용 출력 데이터는 [p, 1-p]일 수 있고, 여기서 p에는 0이 아닌 스코어 값이 부여될 수 있다. 일 실시예에 따르면, 스코어는 환자의 t번째 구간의 생존 기간에 비례하도록 부여될 수 있다. 여기서 생존 기간은, 적어도 월 단위로 구분될 수 있다. 예컨대, 1년 3개월 생존한 환자 D의 구간 별 생존 기간에 따른 스코어는 아래의 [표 4]와 같게 된다.

[130] [표4]

| N (년) | 1 | 2 | 3 | 4 | 5 |
|---------|---|------|---|---|---|
| 구간별 스코어 | 1 | 3/12 | 0 | 0 | 0 |

[131]

[132] 따라서 이 경우 2년 후 생존율을 예측하는 인공신경망을 학습시킬 때, 출력층의 [생존율 노드, 사망률 노드]의 값과 비교될 학습용 출력 데이터는 [3/12, 1-3/12] = [0.75, 0.25]가 될 수 있다.

[133] 이와 같은 방법에 따르면, 사망 등의 이유로 추적 기간이 5년 미만인 환자의 데이터(중도 절단 데이터, right-censored case)의 경우라도 생존율이 0으로 카운트되지 않고 생존 기간만큼의 랭킹화된 스코어가 부여됨으로써, 생존율 예측 모델 생성에 사용되는 유의미한 데이터 수를 늘릴 수 있어, 결과적으로 생존율 예측의 정확도가 향상된다.

[134] 인공신경망은 이러한 학습용 입력 데이터 및 스코어를 이용한 학습용 출력 데이터를 이용하여, 위암 환자의 2년 후 생존율을 예측할 수 있도록 다시 학습될 수 있다.

[135] 이와 같은 과정이 반복됨에 따라($t=t+1$), t년 후 임상 데이터(X_t) 및 N-1년 후 생존율 예측 결과(S_{t-1})를 입력하였을 때 t년 후 생존율(S_t)을 출력할 수 있는 입출력함수인 t년 후 생존율 예측 모델(SM_t)이 인공신경망을 학습시켜 생성된다.

[136] 본 발명의 일 실시예에 따르면, 't-1년 후 시점에서의 환자의 예후'를 반영하는 t-1년 후 생존율 예측 결과(S_{t-1})를 이용하여 t년 후의 생존율을 예측하므로, 각 연차별로 인공신경망을 학습시킬 때마다 생존율 예측 성능이 좋아지게 된다.

[137] 본 발명의 일 실시예에서는, LSTM 알고리즘을 이용하여 생존율 예측 모델을

생성하였다. 이때 연속적인(discrete) 시간 t에서의 생존율(survival probability)은 아래 <수학식 1>과 같이 정해지며, 위험 함수(hazard ration function)는 <수학식 2>와 같이 정해질 수 있다.

[138] <수학식 1>

[139]
$$ft(X) = \sigma(Wt * Xt) = \frac{1}{1 + e^{-\omega_t X_t}}$$

[140] <수학식 2>

[141]
$$HRt(X) = e^{Wt * Xt} = e^{\theta_1 X + \theta_2 t}$$

[142] 이때, 입력 변수 X_t 및 생존 데이터 Y_t 를 가지는 환자 그룹에서, 첫번째 시점 t에서 매개 변수 벡터 $W_t = \theta_1 X + \theta_2 t$ 가 최적화되면 LSTM 층은 W_t 를 기억한다. 이후 생존율 모델은 다음 시점 t+1에서, 생존 데이터 Y_{t+1} 에 대한 매개 변수 벡터 W_{t+1} 을 산출하도록 다시 트레이닝된다. 예를 들어, 위암 발병 2년 후 환자가 질병으로 사망한 경우, 생존율 모델은 첫 해에 생존 데이터 $Y_1 = 1$ 을 학습하고 두 번째 해에는 생존 데이터 $Y_2 = 0$ 을 학습한다. 이때 LSTM은 각 생존 데이터 값(Y)에 대해 매개 변수 W를 기억하고 최적화하기 때문에, RNN 기반의 생존율 예측 모델은 특정 시점에서의 생존율을 산출할 수 있다.

[143] 그런데 시간이 지날 때 마다 임상 데이터를 빠짐없이 수집하는 것은 어렵기 때문에, 환자의 임상 데이터 X는 모든 시점에서 업데이트될 수는 없다. 한편, 생존율 예측 모델의 목적은 특히 환자의 위암 발견 시의 정보 (즉, 병원 첫 방문 시의 정보)를 통해 장기적으로 생존율을 예측하는 것이다. 따라서 본 발명자들은, 환자의 임상 데이터는 관측 시간 동안 일정하나, 시간의 흐름에 의존하며(dependent) 특정 시간에서 환자의 상태를 나타내는 잠재 인자(latent feature)가 있는 것을 가정하였다. 이때 잠재인자 및 위험함수는 <수학식 3>, <수학식 4>와 같이 시간 의존적인 값으로 정의된다.

[144] <수학식 3>

[145]
$$\text{시간 의존 생존값(Time dependent life value)} = \theta_2 t + \theta_3 \partial S$$

[146] <수학식 4>

[147]
$$\text{시간 의존 위험함수(Time dependent hazard function)} e^{Wt * Xt} = e^{\theta_1 X + \theta_2 t + \theta_3 \partial S}$$

[148]

[149] 이때 시간 의존 생존값은 환자의 임상 데이터(X)에 포함되어, 최종 입력 데이터 X_t 는 시간에 의존하는 데이터가 된다. 시간 의존 생존값은 시간(t) 정보와 경사 하강 방정식(gradient descent equation) ∂S 에 의해 얻은 생존율 예측값(S_t) 정보를 포함한다.

[150] 한편, 경사 하강 방정식 ∂S 는 아래의 <수학식 5> 내지 <수학식 8>에 의해 정해질 수 있다.

[151] <수학식 5>

[152]
$$S_{t+1} = S_t + \partial S$$

[153] <수학식 6>

[154]
$$S_{t+1} = S_t + \alpha(Y_t - \hat{Y}_t) \frac{ft'(x)}{|ft(x)|} * \hat{Y}_t$$

[155] <수학식 7>

[156]
$$\frac{dft(x)}{dx} = -W * \frac{e^{-wx}}{(1 + e^{-wx})^2} \approx \alpha * ft(x)(1 - ft(x))$$

(α : 스텝 크기, $\hat{Y}_t = ft(x)$)

[157] <수학식 8>

[158]
$$S_{t+1} = S_t + \alpha(Y_t - \hat{Y}_t)(1 - \hat{Y}_t)\hat{Y}_t$$

[159]

[160] 상술한 알고리즘을 통해 위암 환자의 생존율을 예측하는 모델 생성이 완료되면, 인공신경망의 각 노드의 연결에 대응하는 계수(weight)가 생존율 예측에 최적화되도록 학습된 상태가 된다. 따라서 임의의 위암 환자의 임상 데이터로부터 얻은 입력 데이터를 인공신경망의 입력층에 입력하여 출력층에 출력된 값을 통해 환자의 생존율을 예측할 수 있다. 즉 본 발명에 따른 인공신경망을 이용한 위암의 예후 예측 방법에 따르면, 위암 환자의 예후를 개개인별로 정확하게 예측할 수 있다.

[161]

[162] 도 6는 본 발명의 일 실시예에 따라 연도별로 순차적으로 학습시킨 인공신경망의 히트맵(heatmap) 그래프의 일부를 개략적으로 나타낸 그림이다. 본 발명자들은 일 실시예에서 도 6과 같이 인공신경망을 연도별로 순차적으로 학습시켜 연도별 생존율 예측 모델을 생성한 후, 각각의 모델의 성능을 평가하였다.

[163]

도 7은 트레이닝 중 연도별 생존율 예측 모델의 ROC 그래프의 AUC값을 비교한 그래프이다. 100번의 트레이닝 동안, AUC의 평균은 1년 후 생존율 예측 모델에서 0.79±0.052, 2년 후 모델에서 0.839±0.045, 3년 후 모델에서 0.89±0.049, 4년 후 모델에서 0.915±0.05, 5년 후 모델에서 0.92±0.049이었다.

[164]

도 8은 별도의 테스트 데이터로 연도별 생존율 예측 모델을 검증한 ROC 그래프이다. AUC의 값은 1년 후 생존율 예측 모델에서 0.858, 2년 후 모델에서, 0.869, 3년 후 모델에서 0.879, 4년 후 모델에서 0.912, 5년 후 모델에서 0.923로 나타나, 시간이 지날수록 예측 모델의 성능이 좋아지는 것을 확인할 수 있다.

[165]

도 9는 생존율 예측 모델이 예측한 생존율과 실제 생존 비율을 비교한 그래프이다. 카플란-마이어(Kaplan-Meier) 생존 분석 결과, 생존율 예측 결과와 실제 생존 비율은 95% 신뢰 구간 내에서 15% 오차 범위(점선)로 상관성이 있음을 확인하였다.

[166]

도 10 내지 도 14는 각각 1년, 2년, 3년, 4년, 5년 후 생존율 예측 모델의 판단

곡선(decision curve)을 나타낸 그래프이다. AUC는 단순히 예측의 정확도만을 평가하지만, 판단 곡선은 임상적 결과를 반영하여 임상적 판단의 기준이 되는 임계 확률(threshold probability)에 대한 각각의 순편익(net benefit)을 계산하고 가시화한다. 따라서 판단 곡선을 통해 예측 모델의 실제 임상에서의 가치를 평가할 수 있다. 도 10 내지 도 14을 참조하면 모든 임계 확률에서 순편익이 양의 값을 보이고, 특히 연차가 높아질수록 순편익이 높아지는 것을 확인할 수 있다. 즉 본 발명의 생존율 예측 모델이 임상 판단에 도움이 됨을 알 수 있다.

[167]

[168] 도 15는 본 발명의 일 실시예에 따른 인공신경망과 비교용 단순 인공신경망의 학습 효과를 나타낸 그래프이다. 도 15에서는 RNN층을 포함하는 본 발명의 일 실시예에 따른 인공신경망을 RSN(Recurrent Survival Network)으로, 비교용 단순신경망은 Simple_NN으로 나타내었다.

[169] 도 15의 (a)는 반복 학습(nb_epoch)에 따른 에러(cross_entropy)를 나타낸 그래프이다. 비교용 단순 신경망(Simple_NN)의 경우 에러의 절댓값이 크며 감소율 또한 작으나, 본 발명에 따른 인공신경망(RSN)의 경우 반복 학습에 따른 에러의 감소가 큼을 확인할 수 있다.

[170] 도 15의 (b)는 각 학습 시의 교차 검증(cross validation)을 보여주는 그래프이다. 그래프에서 보듯, 단순 신경망(Simple_NN)의 경우에는 반복 학습을 하여도 검증 손실(validation loss)의 감소가 적고, 편차가 큰 것으로 보아 모델의 안정성이 떨어진다. 반면, RSN의 경우 검증 손실이 매우 낮고 안정적이다.

[171] 도 15의 (c)는 생존율 예측의 정확도를 보여주는 ROC(receiver operating characteristic) 그래프이며, 도 15의 (d)는 ROC 그래프의 AUC(area under curve) 분포 결과를 나타낸 그래프이다. 생존율 예측의 정확도는 ROC 그래프 아래의 면적인 AUC로 정량화가 가능하며 면적이 1에 가까울수록 정확도가 높다. 본 발명자들이 100회 반복 시행한 결과, RSN의 경우에는 평균 0.95 이상의 정확도를 보였으며, 단순 신경망(Simple_NN)의 정확도는 약 0.70이었다. MedCalc program을 이용한 Mann_Whitney test를 통하여 통계적 검증을 시행한 결과 $p < 0.001$ 로 RSN의 진단 정확도가 단순 신경망(Simple_NN)에 비하여 유의하게 높았다.

[172]

[173] <실시예 2>

[174] 도 16은 타 지역 데이터를 이용하여 인공신경망을 재학습시키는 방법에 대한 모식도이다.

[175] 다른 임상 변수의 데이터베이스를 가진 타 지역 또는 타 병원의 데이터를 이용해 RSN 인공신경망을 재학습시킬 경우, 그 지역 데이터에 최적화된 인공신경망을 새롭게 구성할 수 있다. 이때 임상 데이터가 부족하거나 임상 데이터의 종류에서 차이가 있는 타 지역 데이터만으로 새로운 인공신경망을 학습시키는 것 보다는, 이미 학습된 RSN 모델에 타 지역 데이터를 추가시켜 RSN

모델을 재학습시킬 경우 생존 예측의 정확도가 더 높게 된다.

[176] 본 연구에는 타 지역 데이터로 the Gastric Cancer Project Singapore cohort로부터 얻은 데이터(이하, '싱가포르 데이터')가 이용되었다. 싱가포르 데이터는 사용 가능한 변수의 개수가 12개로 (molecular subtype, sex, age, pstage, peritoneal cytology, met site, p_node, lauren, pathology type, Lymphovascular invasion, recurrence, 5_FU adjuvant), [표 1]의 19개의 변수를 포함하는 RSN 모델을 이용하기에는 부족한 정보를 가지고 있었다.

[177] 본 발명자들은 싱가포르 데이터를 학습용 데이터와 테스트 데이터로 나누고, 싱가포르 학습용 데이터를 기존의 데이터에 추가하여 RSN으로 재학습시킨 인공신경망 모델과 싱가포르 데이터만으로 학습된 로컬 인공신경망 모델의 성능을 비교해 보았다.

[178]

[179] 도 17은 RSN에 싱가포르 데이터를 추가하여 재학습시킨 모델과 싱가포르 데이터만으로 학습된 모델을 비교한 히트맵 그래프이다.

[180] 도 17의 (a)를 참조하면, 싱가포르 데이터만으로 학습된 변수의 개수는 12개이며, 이의 분류를 이용해 입력층에 29개의 노드를 생성하였다. 도 8의 (b)의 경우에는, 싱가포르 데이터의 일부를 학습용 RSN의 학습용 데이터에 추가시켜 20개 변수, 53개 노드를 생성하여 인공신경망을 재학습시켰다. 이때 본래 데이터에는 있고 싱가포르 데이터에는 없는 데이터(missing value, NaN)는 전술한 knn 알고리즘을 통해 추가되었다.

[181]

[182] 도 18는 원래의 모델, 싱가포르 데이터를 추가하여 재학습시킨 재학습 모델, 싱가포르 데이터만으로 학습시킨 싱가포르 모델의 효과를 비교한 그래프이다.

[183] 원래의 모델(original)은 기존의 데이터를 이용하여 기존의 데이터로 테스트한 인공신경망을 의미하고, RSN으로 재학습시킨 모델(adaptive training set)은 기존의 데이터와 싱가포르 데이터를 통해 다시 재학습시키고 싱가포르 데이터로 테스트한 인공신경망을 의미하여, 싱가포르 모델(Singapore training set)은 싱가포르 데이터만을 이용하여 학습시키고 테스트한 인공신경망을 의미한다.

[184] 도 18의 (a)는 반복 학습에 따른 에러의 감소를 나타내고 있다. 그림에서와 같이 재학습 모델의 경우, 반복 학습에 따른 에러의 감소 폭이 큰 것을 보아 학습 속도가 빠른 것을 확인할 수 있다. 반면, 싱가포르 모델의 경우 반복 학습에 따른 에러의 감소가 적다.

[185] 도 18의 (b)는 각 학습 시의 교차 검증 결과의 에러 감소를 보여주는 그래프이다. 이 그래프에서 보듯이 싱가포르 모델의 경우 반복 학습에도 validation loss의 감소가 적고, 편차가 큰 것으로 보아 모델의 안정성이 떨어지나, 재학습 모델의 경우 validation loss가 매우 낮고 안정적이다.

[186] 한편, 도 18의 (a) 및 (b)에서, 재학습 모델은 이미 학습된 인공신경망을

이용하여 재학습하기 때문에, 처음부터 학습을 시작하는 원래의 모델보다도 학습 속도가 빠르다.

- [187] 도 18의 (c)는 예측의 정확도를 보여주는 ROC 그래프이다. 본 발명자들이 100회 반복 시행한 결과 재학습 모델의 경우 싱가포르의 테스트 데이터의 생존율을 평균 0.95의 확률로 맞추었으나, 싱가포르 모델의 경우 정확도가 0.80 내외였다. 맨-휘트니 검정(Mann Whitney test)을 통하여 통계적 검증을 시행한 결과 $p < 0.001$ 로 재학습 모델의 생존율 예측 정확도가 싱가포르 모델보다 유의하게 높았다. 재학습 모델은 원래의 모델의 정확도에 비해서도 크게 낮지 않은 성능을 보여주었다. 즉 이때 knn 알고리즘에 의해 결측치를 추가시키더라도 예측 정확도가 크게 낮아지지 않음을 확인할 수 있다.

[188]

- [189] 도 19는 본 발명의 일 실시예에 따른 인공신경망을 이용한 위암의 예후 예측 장치의 구성을 개략적으로 나타낸 그림이다.

- [190] 도 19에 도시된 위암의 예후 예측 장치(10)는 본 실시예의 특징이 흐려지는 것을 방지하기 위하여 본 실시예와 관련된 구성요소들만을 도시한 것이다. 따라서, 도 19에 도시된 구성요소들 외에 다른 범용적인 구성요소들이 더 포함될 수 있음을 본 실시예와 관련된 기술분야에서 통상의 지식을 가진 자라면 이해할 수 있다.

- [191] 본 발명의 일 실시예에 따른 위암의 예후 예측 장치(10)는 적어도 하나 이상의 프로세서(processor)에 해당하거나, 적어도 하나 이상의 프로세서를 포함할 수 있다. 이에 따라, 위암의 예후 예측 장치(10)는 마이크로프로세서나 범용 컴퓨터 시스템과 같은 다른 하드웨어 장치에 포함된 형태로 구동될 수 있다.

- [192] 본 발명은 기능적인 블록 구성들 및 다양한 처리 단계들로 나타내어질 수 있다. 이러한 기능 블록들은 특정 기능들을 실행하는 다양한 개수의 하드웨어 또는/및 소프트웨어 구성들로 구현될 수 있다. 예를 들어, 본 발명은 하나 이상의 마이크로프로세서들의 제어 또는 다른 제어 장치들에 의해서 다양한 기능들을 실행할 수 있는, 메모리, 프로세싱, 로직(logic), 룩업 테이블(look-up table) 등과 같은 직접 회로 구성들을 채용할 수 있다. 본 발명에의 구성 요소들이 소프트웨어 프로그래밍 또는 소프트웨어 요소들로 실행될 수 있는 것과 유사하게, 본 발명은 데이터 구조, 프로세스들, 루틴들 또는 다른 프로그래밍 구성들의 조합으로 구현되는 다양한 알고리즘을 포함하여, C, C++, 자바(Java), 어셈블러(assembly) 등과 같은 프로그래밍 또는 스크립팅 언어로 구현될 수 있다. 기능적인 측면들은 하나 이상의 프로세서들에서 실행되는 알고리즘으로 구현될 수 있다. 또한, 본 발명은 전자적인 환경 설정, 신호 처리, 및/또는 데이터 처리 등을 위하여 종래 기술을 채용할 수 있다. "매커니즘", "요소", "수단", "구성"과 같은 용어는 넓게 사용될 수 있으며, 본 발명의 구성요소들이 기계적이고 물리적인 구성들로서 한정되는 것은 아니다. 상기 용어는 프로세서 등과 연계하여 소프트웨어의 일련의 처리들(routines)의 의미를 포함할 수 있다.

- [193] 도 19를 참조하면, 위암의 예후 예측 장치(10)는 데이터 획득부(11), 인공신경망 학습부(12) 및 생존율 예측 모델 생성부(13)를 포함한다.
- [194] 데이터 획득부(11)는 복수의 위암 환자들의 의료 데이터, 예컨대 임상 데이터를 획득 및 위암 발병 후 생존 기간 데이터를 획득한다. 임상 데이터는 환자의 의료 영상으로부터 획득되거나, 환자의 검체 검사 결과로부터 획득될 수 있으나, 이에 한정되지 않는다.
- [195] 인공신경망 학습부(12)는 복수의 위암 환자들의 임상 데이터와 생존 기간 데이터로부터 학습용 입력 데이터와 학습용 출력 데이터를 획득하고, 학습용 입력 데이터와 학습용 출력 데이터를 이용하여 입력층, 은닉층, 출력층을 포함하는 인공신경망을 학습시킨다.
- [196] 생존율 예측 모델 생성부(13)는 학습된 인공신경망을 이용하여 위암 환자의 생존율을 예측하는 모델을 생성한다. 이때 생존율을 예측한다는 것은 위암 환자의 임상 정보를 입력하면 소정의 알고리즘을 통해 상기 환자의 생존율을 산출한다는 것을 의미할 수 있다.
- [197] 일 실시예에 따르면, 학습용 입력 데이터는 복수의 위암 환자들의 분자 유전학적 아형(subtype) 데이터를 포함할 수 있고, 상기 아형은 MSI(microsatellite instable) 아형, MSS/EMT 아형, MSS/TP53+ 아형, MSS/TP53- 아형을 포함한다. 이때 입력층은 분자 유전학적 아형 데이터가 입력되는 4개의 노드를 포함할 수 있다.
- [198] 일 실시예에 따르면, 인공신경망 학습부(12)는 학습용 입력 데이터의 각 변수를 2차원 이상의 벡터로 임베딩(embedding)하여 임베딩층을 산출할 수 있다.
- [199] 일 실시예에 따르면, 인공신경망 학습부(12)는, k-최근접 이웃 알고리즘(k-nearest neighbor algorithm: knn)을 이용하여 학습용 입력 데이터의 결측치(missing data, NaN)를 추가할 수 있다.
- [200] 일 실시예에 따르면, 인공신경망의 은닉층은 적어도 하나의 순환신경망(RNN)층을 포함할 수 있다.
- [201] 한편, 도 1에 도시된 본 발명의 일 실시예에 따른 인공신경망을 이용한 위암의 예후 예측 방법은 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 같은 저장매체를 포함한다.
- [202] 본 발명에 따른 인공신경망을 이용한 위암의 예후 예측 방법, 장치 및 프로그램에 따르면, 위암 환자의 예후를 개개인별로 정확하게 예측할 수 있다. 그뿐만 아니라, 학습된 인공신경망을 이용하여 각 치료 방법에 의한 예후를 시뮬레이션할 수 있으므로 환자별 맞춤형 치료 방법을 결정할 수 있다.
- [203] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 당해 기술 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형

및 균등한 다른 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

[204]

산업상 이용가능성

[205] 본 발명은 인공지능망을 이용한 위암의 예후 예측 방법, 장치 및 프로그램에 관한 것으로, 진단 및 치료 장치 산업에 이용될 수 있다.

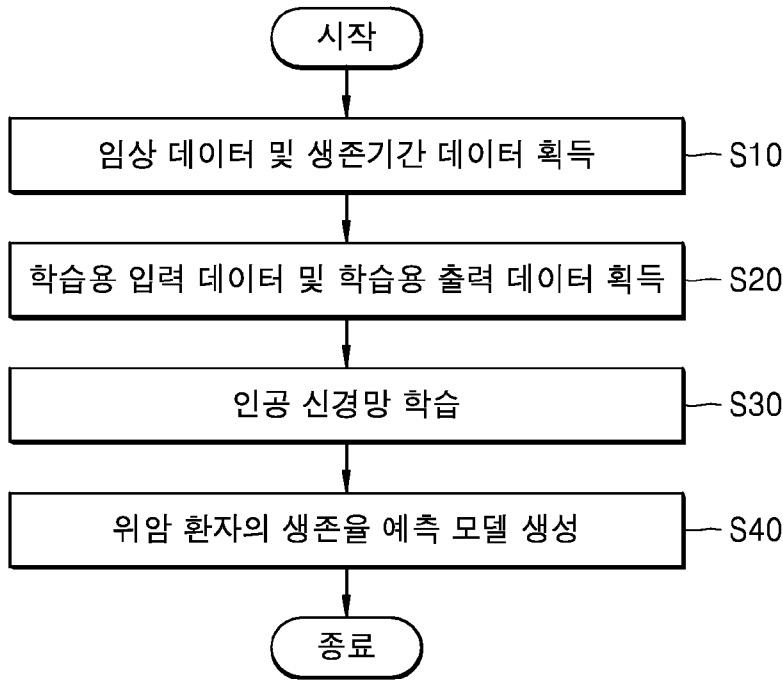
청구범위

- [청구항 1] 복수의 위암 환자들의 임상 데이터 및 위암 발병 후 생존 기간 데이터를 획득하는 단계;
 상기 임상 데이터와 상기 생존 기간 데이터로부터 학습용 입력 데이터와 학습용 출력 데이터를 획득하는 단계;
 상기 학습용 입력 데이터와 상기 학습용 출력 데이터를 이용하여 입력층, 은닉층, 출력층을 포함하는 인공신경망을 학습시키는 단계; 및
 상기 학습된 인공신경망을 이용하여 위암 환자의 생존율을 예측하는 모델을 생성하는 단계;를 포함하는, 인공신경망을 이용한 위암의 예후 예측 방법.
- [청구항 2] 제1항에 있어서,
 상기 학습용 입력 데이터는 상기 복수의 위암 환자들의 분자 유전학적 아형(subtype) 데이터를 포함하고, 상기 아형은 MSI(microsatellite instable) 아형, MSS/EMT 아형, MSS/TP53+ 아형, MSS/TP53- 아형을 포함하는, 인공신경망을 이용한 위암의 예후 예측 방법.
- [청구항 3] 제2항에 있어서,
 상기 입력층은 상기 분자 유전학적 아형 데이터가 입력되는 4개의 노드를 포함하는, 인공신경망을 이용한 위암의 예후 예측 방법.
- [청구항 4] 제1항에 있어서,
 상기 학습용 입력 데이터와 상기 학습용 출력 데이터를 이용하여 상기 인공신경망을 학습시키는 단계는,
 상기 학습용 입력 데이터의 각 변수를 2차원 이상의 벡터로 임베딩(embedding)하여 임베딩층을 산출하는 단계를 포함하는, 인공신경망을 이용한 위암의 예후 예측 방법.
- [청구항 5] 제1항에 있어서,
 상기 임상 데이터와 상기 생존 기간 데이터로부터 각각 학습용 입력 데이터와 학습용 출력 데이터를 획득하는 단계는,
 k-최근접 이웃 알고리즘(k-nearest neighbor algorithm: knn)을 이용하여 결측치(missing data, NaN)를 추가하는 단계를 포함하는, 인공신경망을 이용한 위암의 예후 예측 방법.
- [청구항 6] 제1항에 있어서,
 상기 인공신경망의 상기 은닉층은 적어도 하나의 순환신경망(RNN) 층을 포함하는, 인공신경망을 이용한 위암의 예후 예측 방법.
- [청구항 7] 복수의 위암 환자들의 임상 데이터 및 위암 발병 후 생존 기간 데이터를 획득하는 데이터 획득부;
 상기 임상 데이터와 상기 생존 기간 데이터로부터 학습용 입력 데이터와 학습용 출력 데이터를 획득하고, 상기 학습용 입력 데이터와 상기 학습용

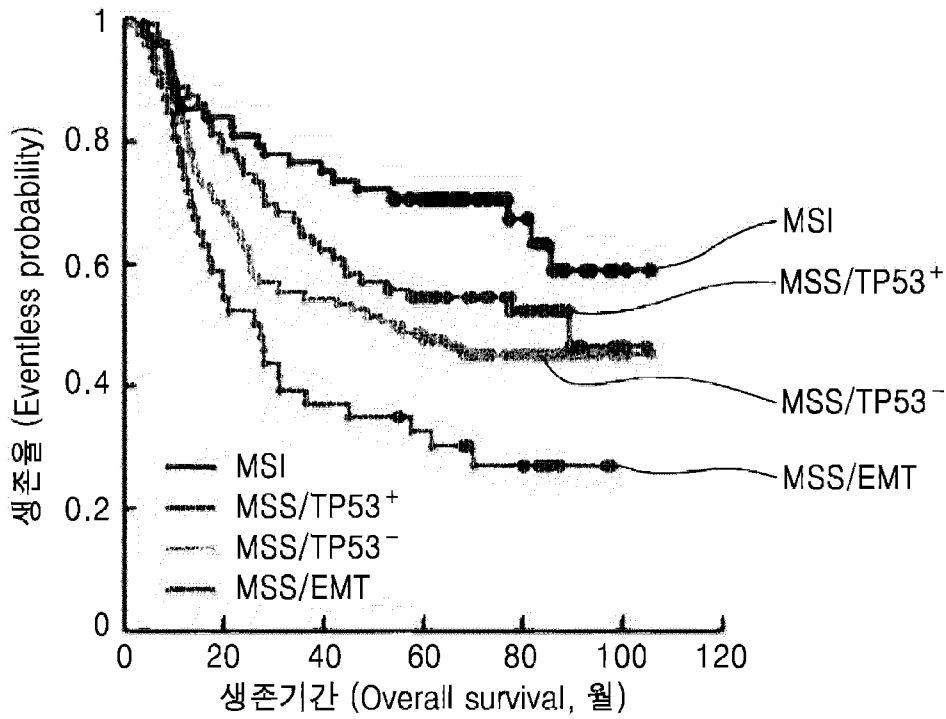
출력 데이터를 이용하여 입력층, 은닉층, 출력층을 포함하는 인공신경망을 학습시키는 인공신경망 학습부; 및 상기 학습된 인공신경망을 이용하여 위암 환자의 생존율을 예측하는 모델을 생성하는 생존율 예측 모델 생성부;를 포함하는, 인공신경망을 이용한 위암의 예후 예측 장치.

- [청구항 8] 제7항에 있어서,
상기 학습용 입력 데이터는 상기 복수의 위암 환자들의 분자 유전학적 아형 데이터를 포함하고, 상기 아형은 MSI(microsatellite instable) 아형, MSS/EMT 아형, MSS/TP53+ 아형, MSS/TP53- 아형을 포함하는, 인공신경망을 이용한 위암의 예후 예측 장치.
- [청구항 9] 제8항에 있어서,
상기 입력층은 상기 분자 유전학적 아형 데이터가 입력되는 4개의 노드를 포함하는, 인공신경망을 이용한 위암의 예후 예측 장치.
- [청구항 10] 제7항에 있어서,
상기 인공신경망 학습부는, 상기 학습용 입력 데이터의 각 변수를 2차원 이상의 벡터로 임베딩(embedding)하여 임베딩층을 산출하는, 인공신경망을 이용한 위암의 예후 예측 장치.
- [청구항 11] 제7항에 있어서,
상기 인공신경망 학습부는, k-최근접 이웃 알고리즘(k-nearest neighbor algorithm: knn)을 이용하여 상기 학습용 입력 데이터의 결측치(missing data, NaN)를 추가하는, 인공신경망을 이용한 위암의 예후 예측 장치.
- [청구항 12] 제7항에 있어서,
상기 인공신경망의 상기 은닉층은 적어도 하나의 순환신경망(RNN) 층을 포함하는, 인공신경망을 이용한 위암의 예후 예측 장치.
- [청구항 13] 컴퓨터를 이용하여 제1항 내지 제6항 중 어느 한 항의 방법을 실행하기 위하여 매체에 저장된 컴퓨터 프로그램.

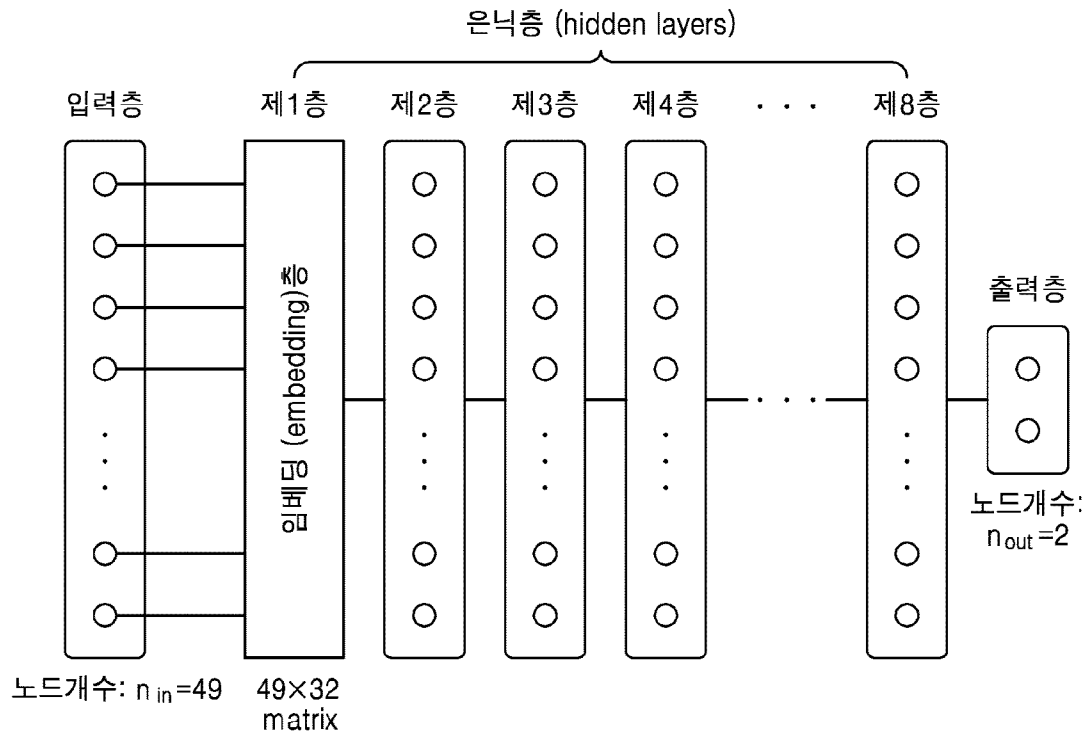
[도1]

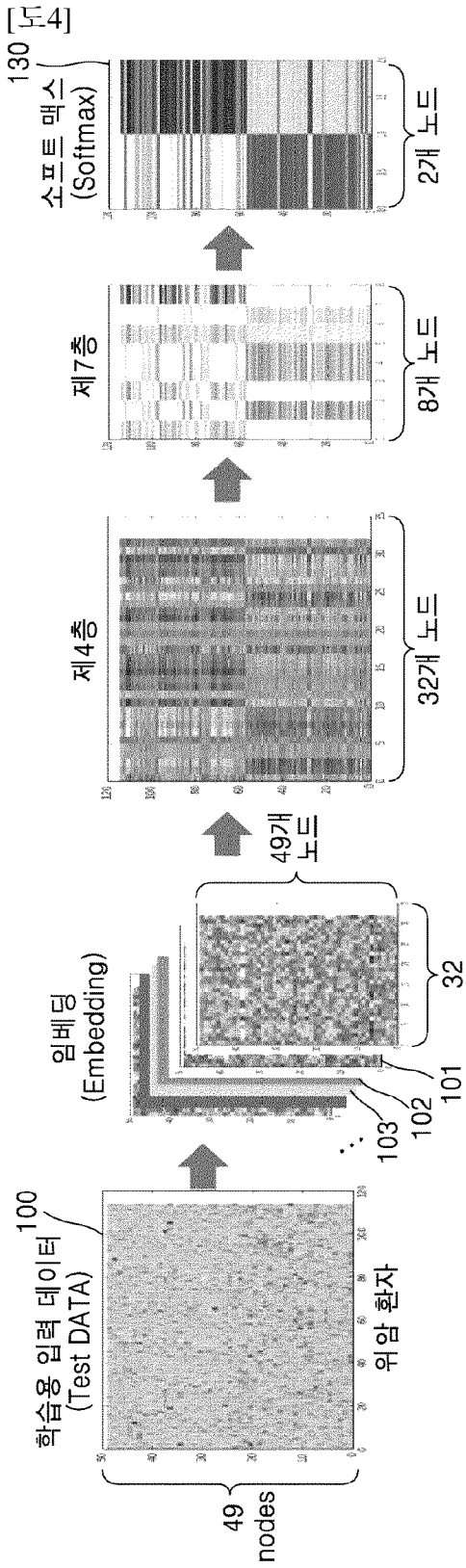


[도2]

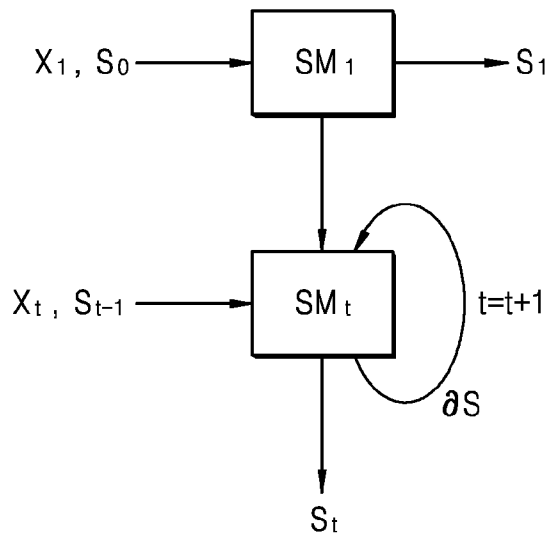


[도3]

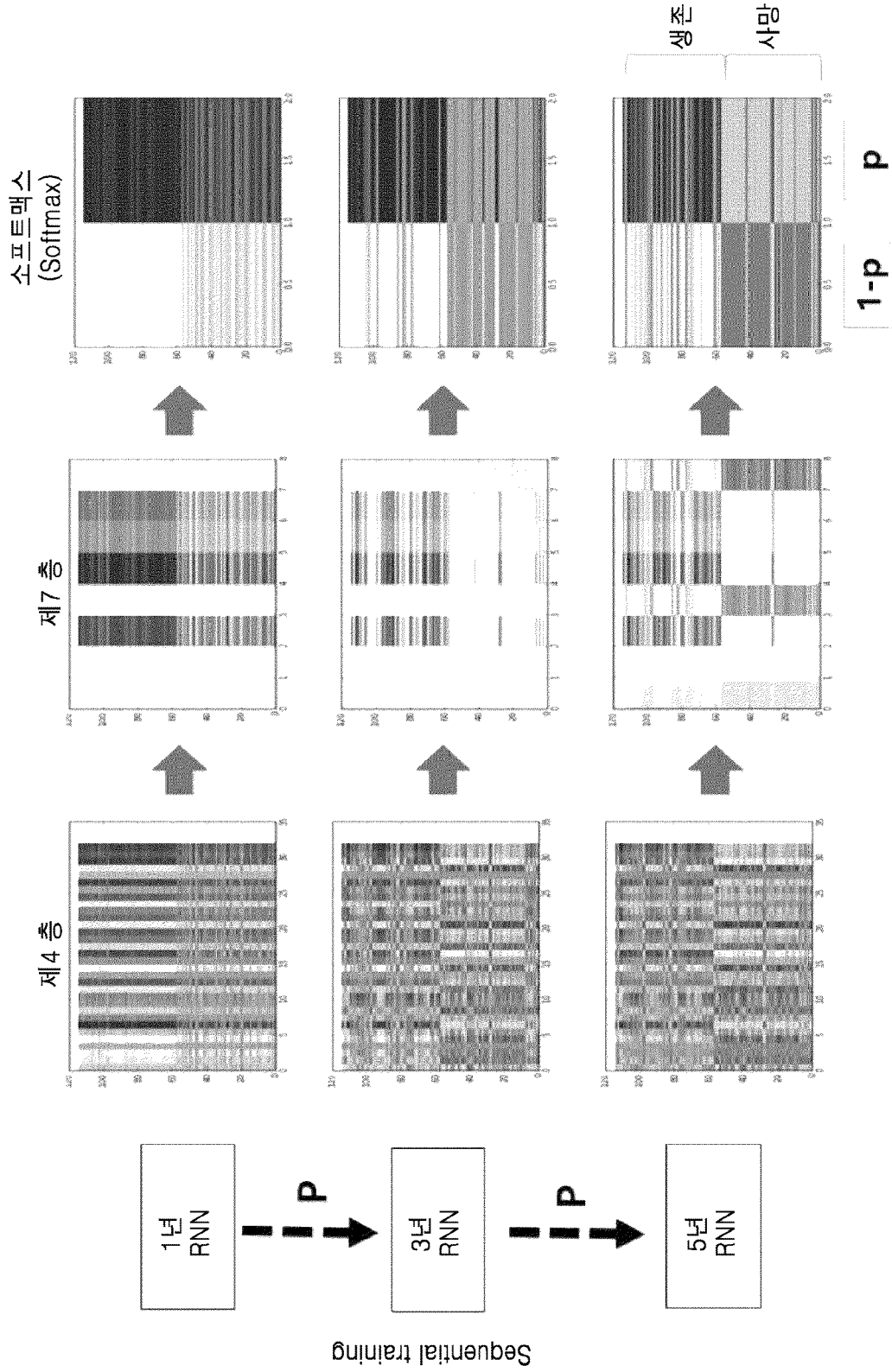




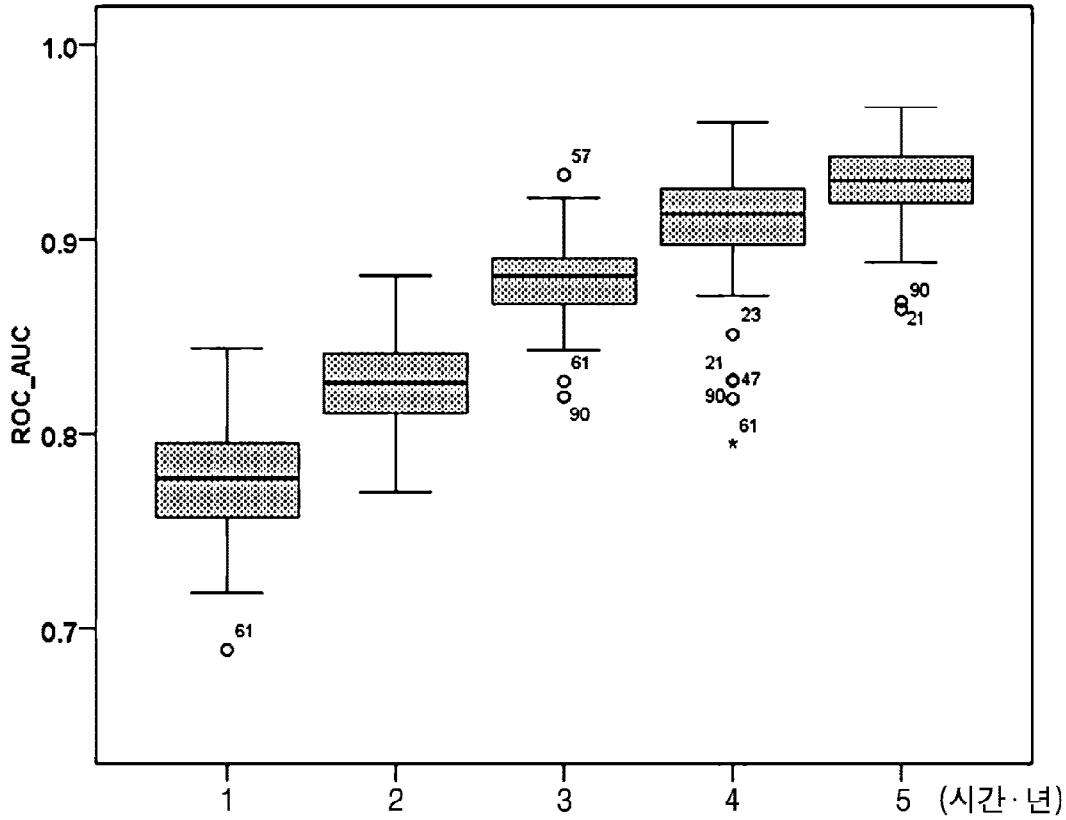
[도5]



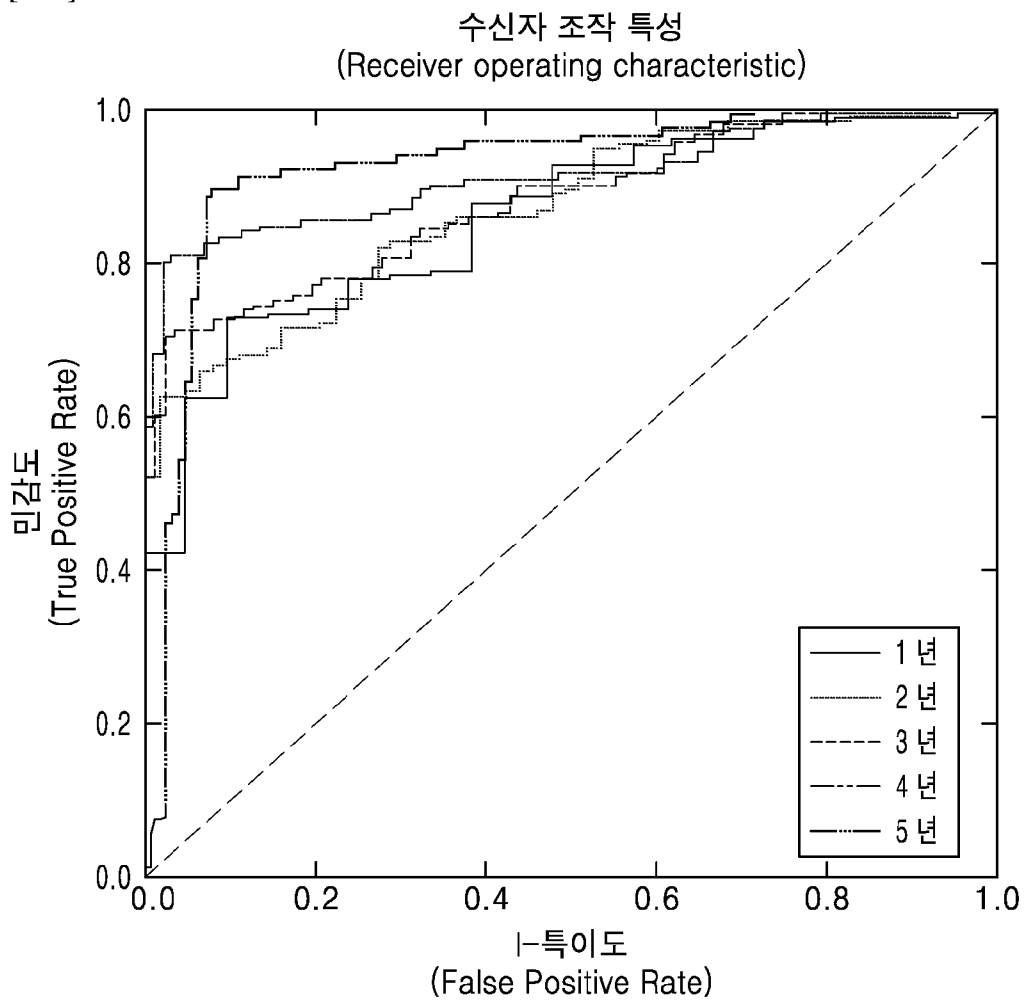
[도6]



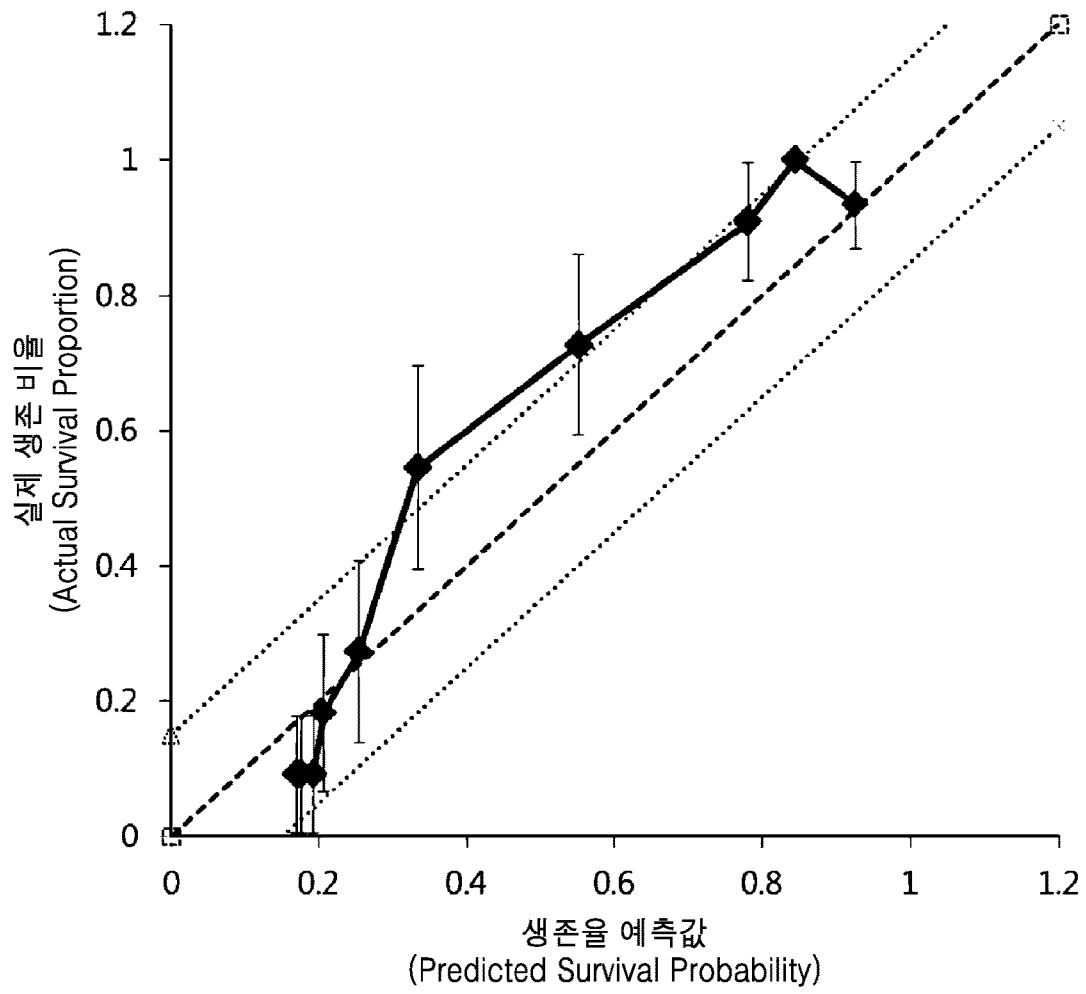
[도7]



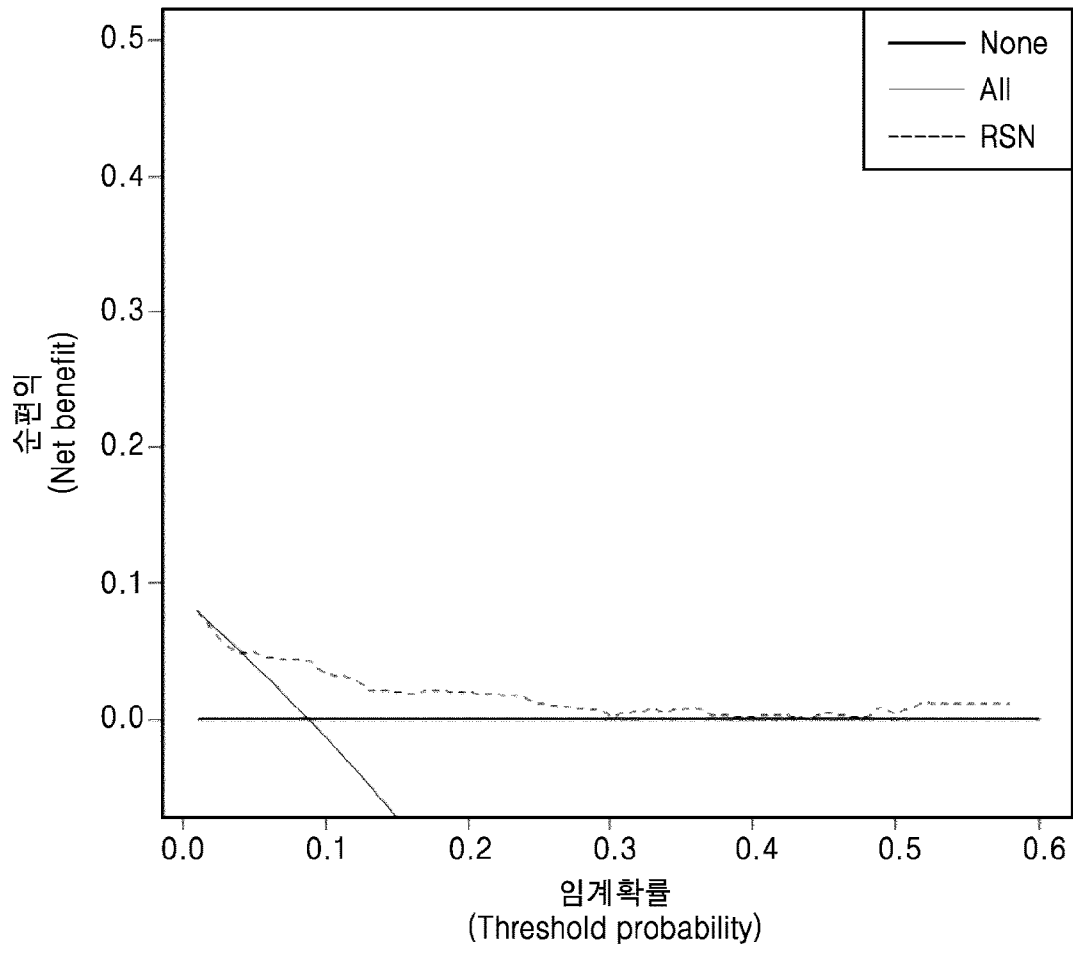
[도8]



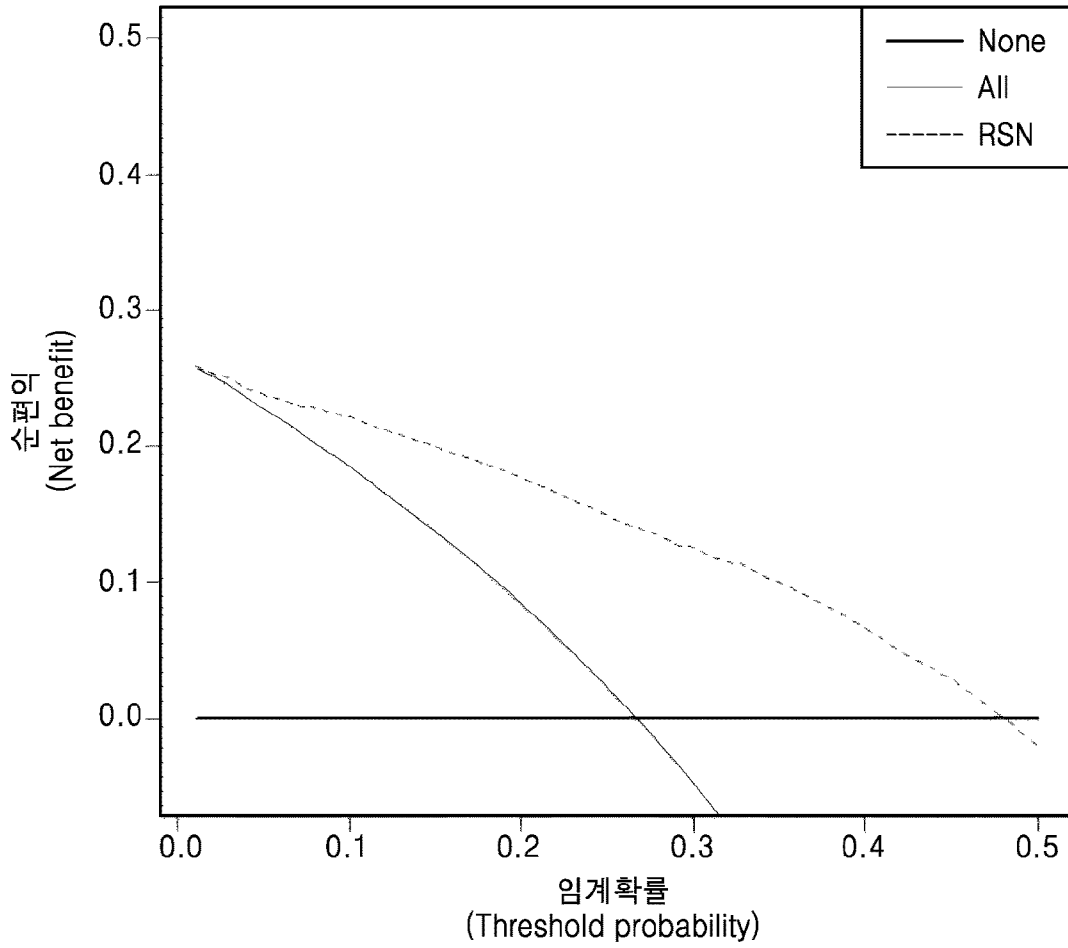
[도9]



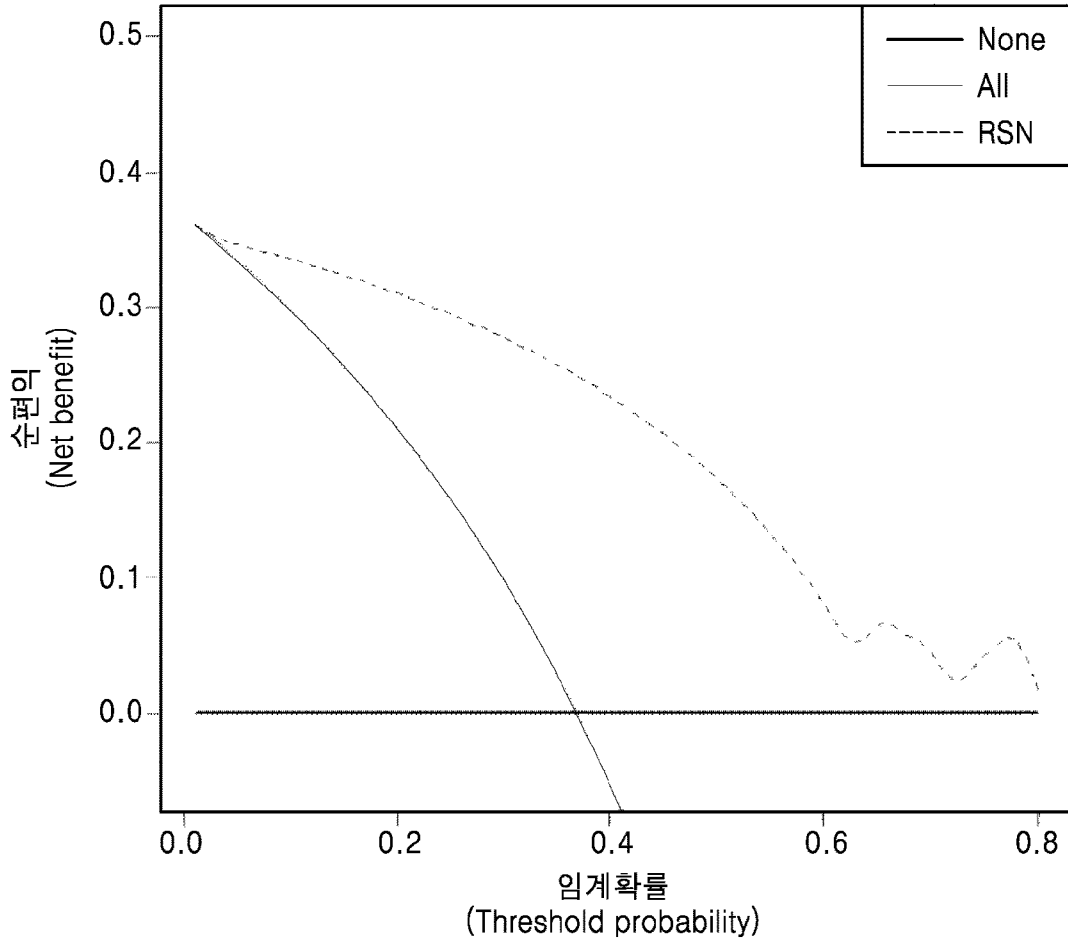
[도10]



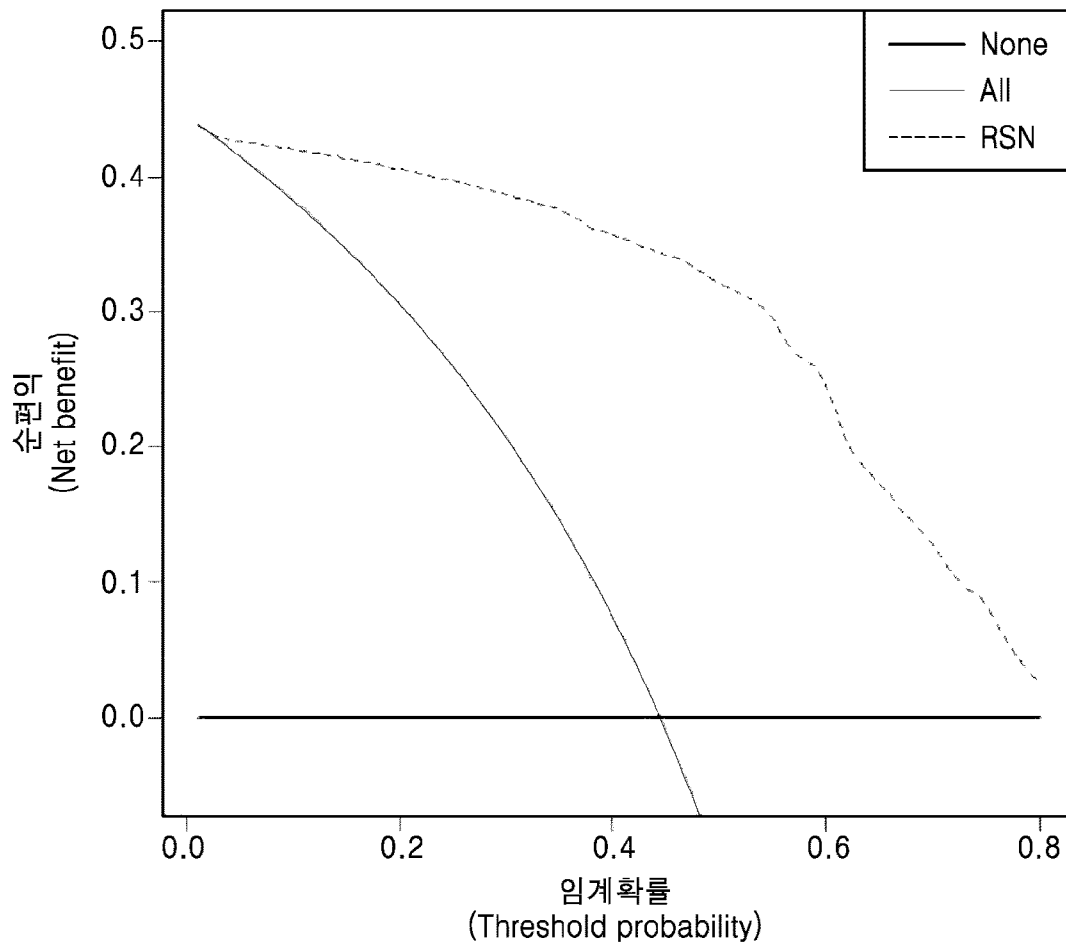
[도11]



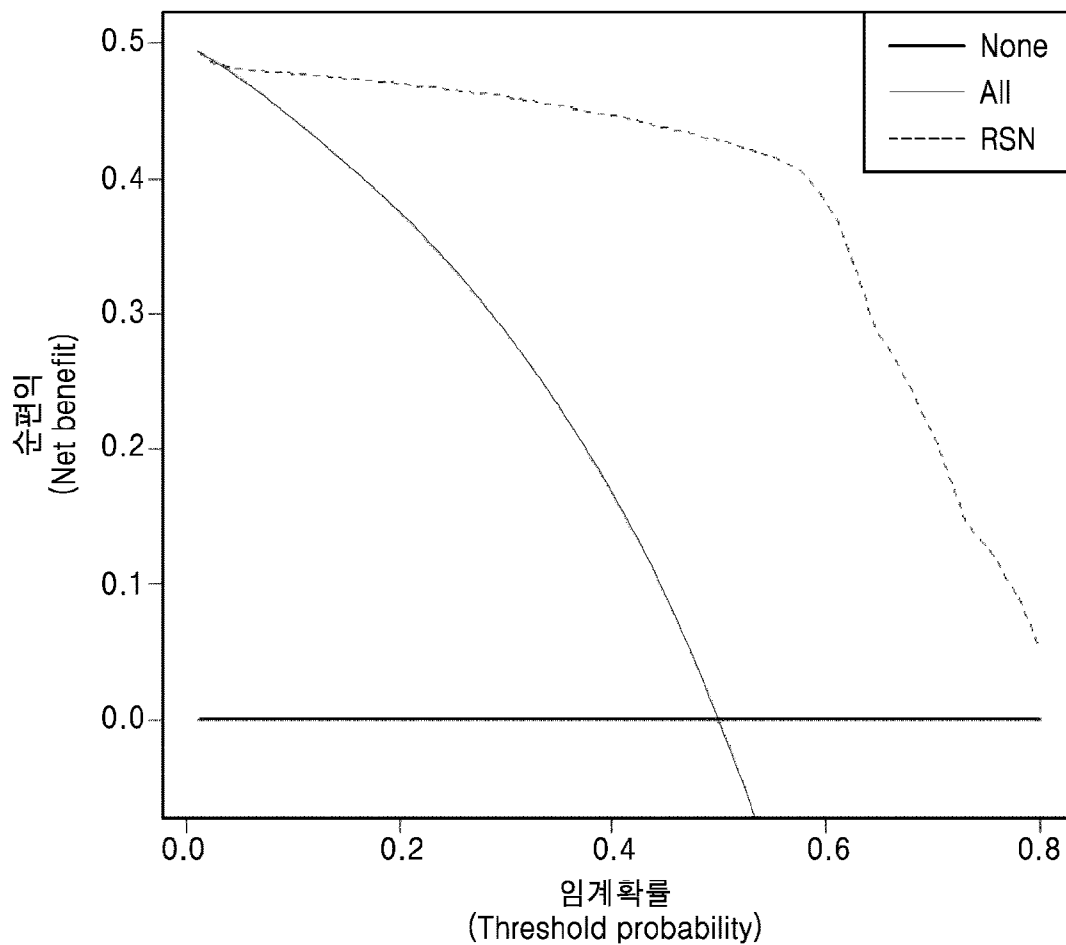
[도12]



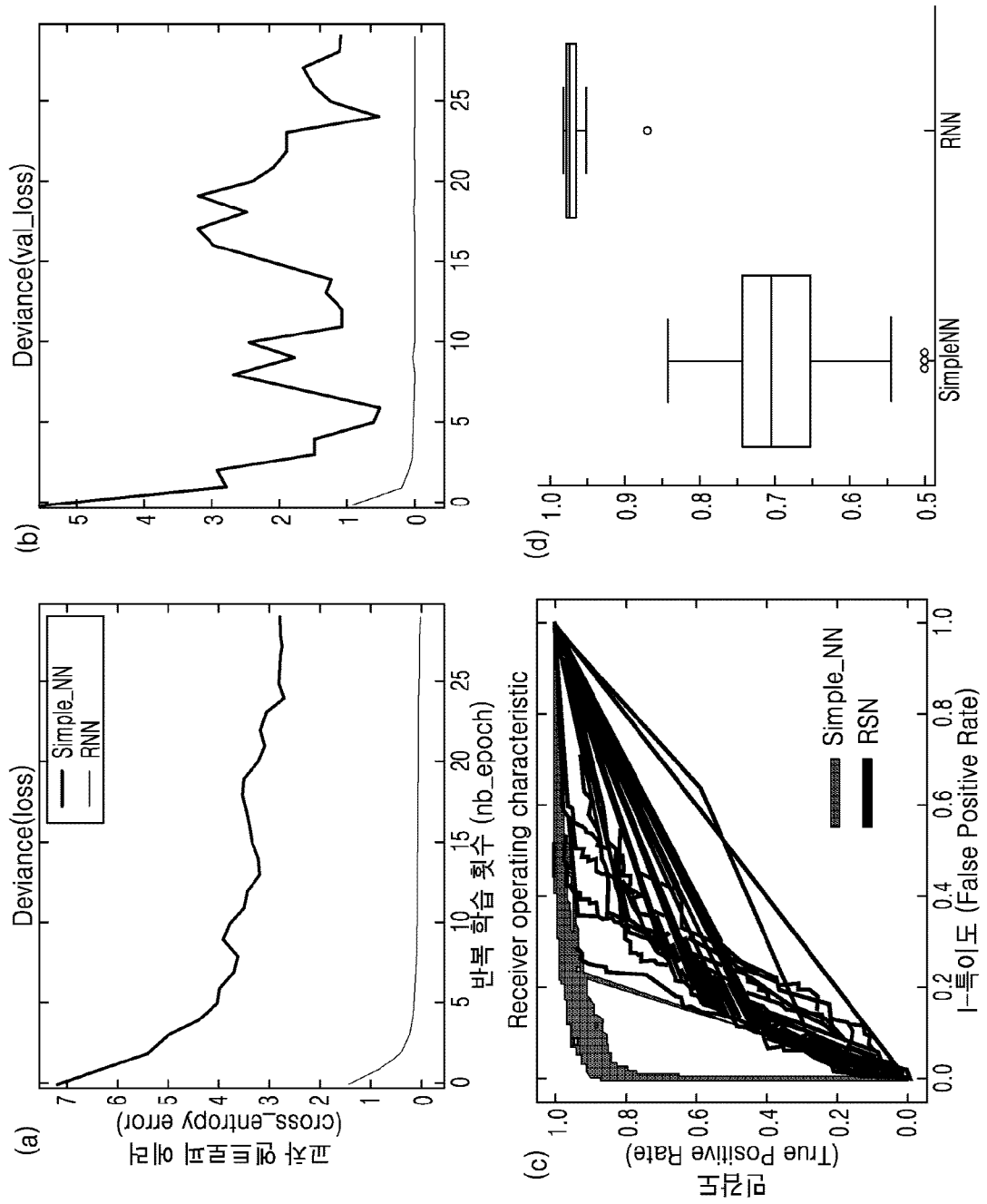
[도13]



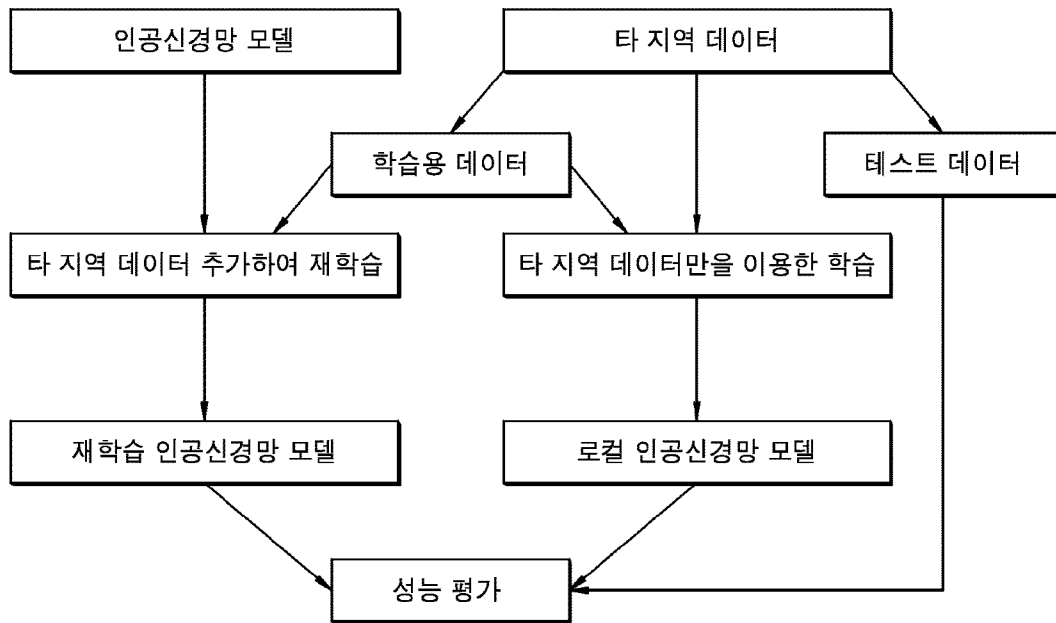
[도14]



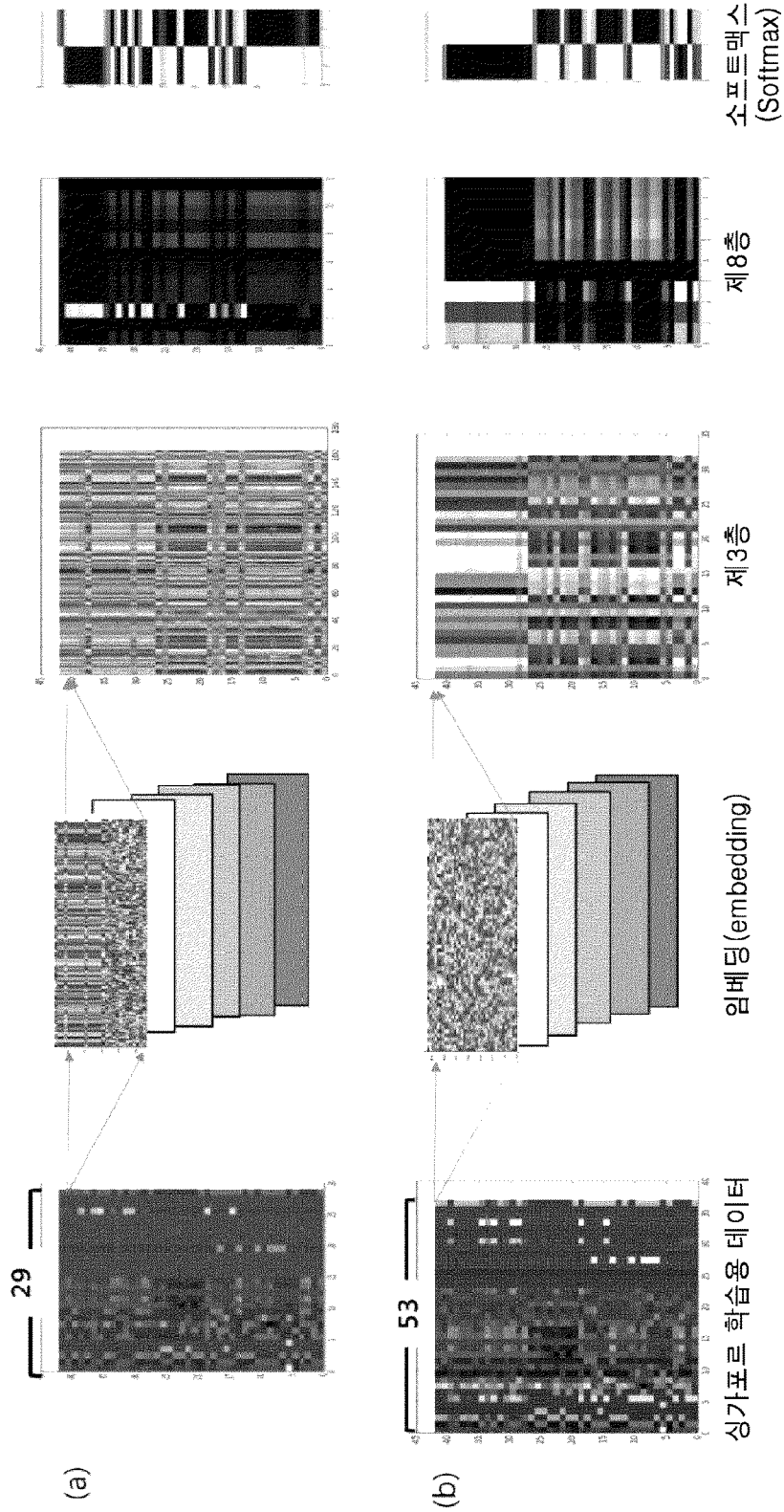
[도 15]



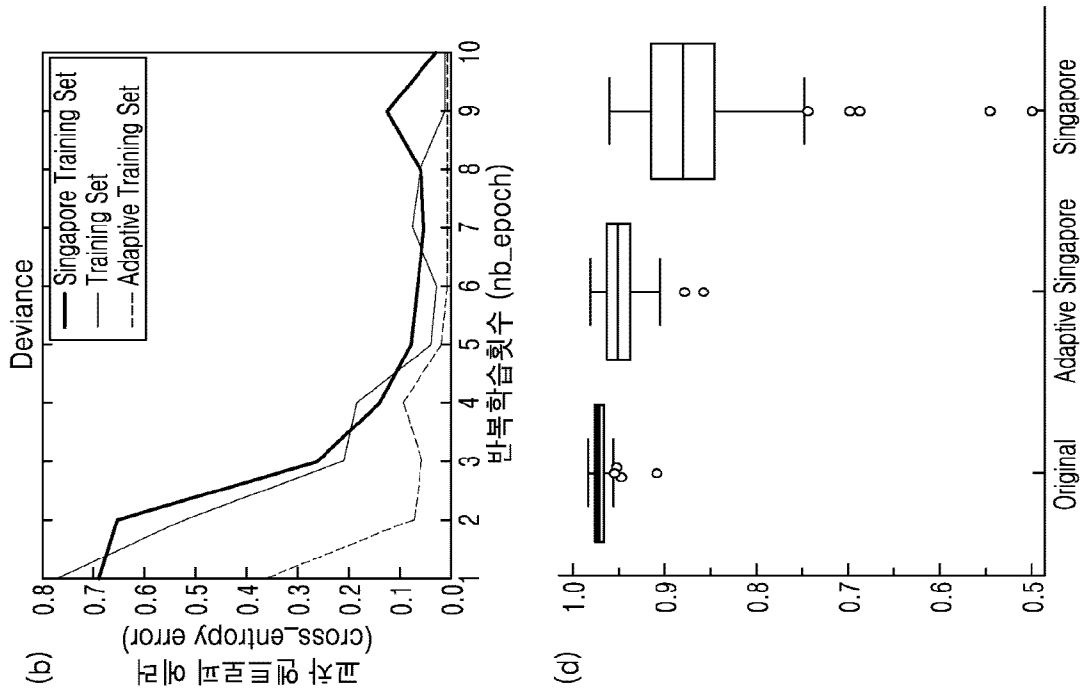
[도16]



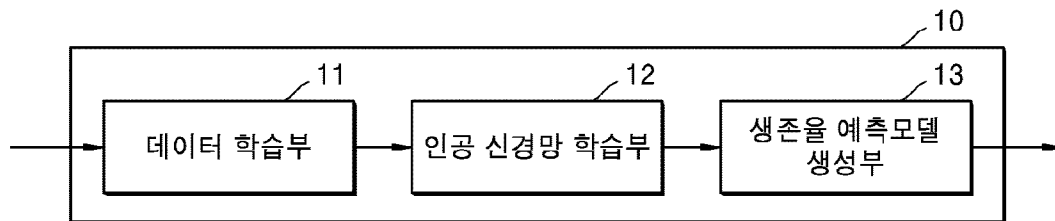
[도17]



[도 18]



[도 19]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2017/012068

A. CLASSIFICATION OF SUBJECT MATTER

G06F 19/00(2011.01)i, G06N 3/08(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 19/00; G06E 1/00; G06N 3/08

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean Utility models and applications for Utility models: IPC as above
Japanese Utility models and applications for Utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal), Google Search Engine & Keywords: gastric cancer, survival period, artificial neural network(ANN), learning, input layer, hidden layer, output layer, prediction, embedding layer, subtype

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|---|-----------------------|
| X | US 2006-0195269 A1 (YEATMAN, Timothy J. et al.) 31 August 2006 See paragraphs [0056], [0091]-[0092], [0097], [0138]; and claims 1, 8-9. | 1,5-7,11-13 |
| Y | | 2-3,8-9 |
| A | | 4,10 |
| Y | CRISTESCU, Razvan et al., "Molecular Analysis of Gastric Cancer Identifies Subtypes Associated with Distinct Clinical Outcomes", Nature Medicine, vol. 21, no. 5, May 2015, pages 449-456. <URL: http://www.nature.com/articles/nm.3850 > See page 450; and figure 1. | 2-3,8-9 |
| A | US 2004-0010481 A1 (MANI, D. R. et al.) 15 January 2004 See paragraphs [0010], [0025]-[0034], [0039]; claim 1; and figures 1, 3. | 1-13 |
| A | WO 2004-015608 A2 (PHASE-IT INTELLIGENT SOLUTIONS AG.) 19 February 2004 See pages 3, 8-17; claims 1-3, 9; and figure 1. | 1-13 |
| A | JANG, Kyung Hwan et al., "Comparison of Survival Prediction of Rats with Hemorrhagic Shocks Using Artificial Neural Network and Support Vector Machine", Journal of the Institute of Electronics Engineers of Korea, vol. 48, SC edition, no. 2, March 2011, pages 47-55. <URL: http://www.dbpia.co.kr/Article/NODE01622360 > See pages 47-50; and figure 2. | 1-13 |



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

23 FEBRUARY 2018 (23.02.2018)

Date of mailing of the international search report

23 FEBRUARY 2018 (23.02.2018)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 189 Seonsa-ro, Daejeon 302-701,
Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2017/012068

| Patent document cited in search report | Publication date | Patent family member | Publication date |
|--|------------------|--|--|
| US 2006-0195269 A1 | 31/08/2006 | EP 1894131 A2 EP 1894131 B1 US 2006-0195266 A1 WO 2006-093507 A2 WO 2006-093507 A3 | 05/03/2008 23/01/2013 31/08/2006 08/09/2006 04/06/2009 |
| US 2004-0010481 A1 | 15/01/2004 | NONE | |
| WO 2004-015608 A2 | 19/02/2004 | AU 2003-260356 A1 EP 1394715 A1 WO 2004-015608 A3 | 25/02/2004 03/03/2004 23/12/2004 |

A. 발명이 속하는 기술분류(국제특허분류(IPC))
G06F 19/00(2011.01)i, G06N 3/08(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)
G06F 19/00; G06E 1/00; G06N 3/08

조사된 기술분야에 속하는 최소문헌 이외의 문헌
한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC
일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))
eKOMPASS(특허청 내부 검색시스템), 구글 검색 엔진 & 키워드: 위암, 생존 기간, 인공 신경망(ANN), 학습, 입력층, 은닉층, 출력층, 예측, 임베딩층, 아형

C. 관련 문헌

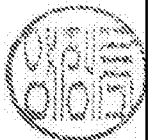
| 카테고리* | 인용문헌명 및 관련 구절(해당하는 경우)의 기재 | 관련 청구항 |
|-------|---|-------------|
| X | US 2006-0195269 A1 (TIMOTHY J. YEATMAN 등) 2006.08.31 단락 [0056], [0091]-[0092], [0097], [0138]; 및 청구항 1, 8-9 참조. | 1,5-7,11-13 |
| Y | | 2-3,8-9 |
| A | | 4,10 |
| Y | RAZVAN CRISTESCU 등, `Molecular Analysis of Gastric Cancer Identifies Subtypes Associated with Distinct Clinical Outcomes`, 네이처 메디슨, 21권, 5호, 2015. 05, 페이지 449-456. <URL: http://www.nature.com/articles/nm.3850 > 페이지 450; 및 도면 1 참조. | 2-3,8-9 |
| A | US 2004-0010481 A1 (D. R. MANI 등) 2004.01.15 단락 [0010], [0025]-[0034], [0039]; 청구항 1; 및 도면 1, 3 참조. | 1-13 |
| A | WO 2004-015608 A2 (PHASE-IT INTELLIGENT SOLUTIONS AG) 2004.02.19 페이지 3, 8-17; 청구항 1-3, 9; 및 도면 1 참조. | 1-13 |
| A | 장정환 등, `출혈성 쇼크를 일으킨 흰쥐에서 인공신경망과 지원벡터기체를 이용한 생존을 비교`, 전자공학회논문지 제48권 SC편 제2호, 2011.03, 페이지 47-55. <URL: http://www.dbpia.co.kr/Article/NODE01622360 > 페이지 47-50; 및 도면 2 참조. | 1-13 |

추가 문헌이 C(계속)에 기재되어 있습니다. 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:
 “A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌
 “E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌
 “L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌
 “O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌
 “P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌
 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌
 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.
 “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.
 “&” 동일한 대응특허문헌에 속하는 문헌

| | |
|--|---|
| 국제조사의 실제 완료일 2018년 02월 23일 (23.02.2018) | 국제조사보고서 발송일 2018년 02월 23일 (23.02.2018) |
|--|---|

| | |
|---|------------------------------------|
| ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578 | 심사관 양정록 전화번호 +82-42-481-5709 |
|---|------------------------------------|



| 국제조사보고서에서 인용된 특허문헌 | 공개일 | 대응특허문헌 | 공개일 |
|-----------------------|------------|--|--|
| US 2006-0195269 A1 | 2006/08/31 | EP 1894131 A2 EP 1894131 B1 US 2006-0195266 A1 WO 2006-093507 A2 WO 2006-093507 A3 | 2008/03/05 2013/01/23 2006/08/31 2006/09/08 2009/06/04 |
| US 2004-0010481 A1 | 2004/01/15 | 없음 | |
| WO 2004-015608 A2 | 2004/02/19 | AU 2003-260356 A1 EP 1394715 A1 WO 2004-015608 A3 | 2004/02/25 2004/03/03 2004/12/23 |