

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2024년 8월 2일 (02.08.2024)



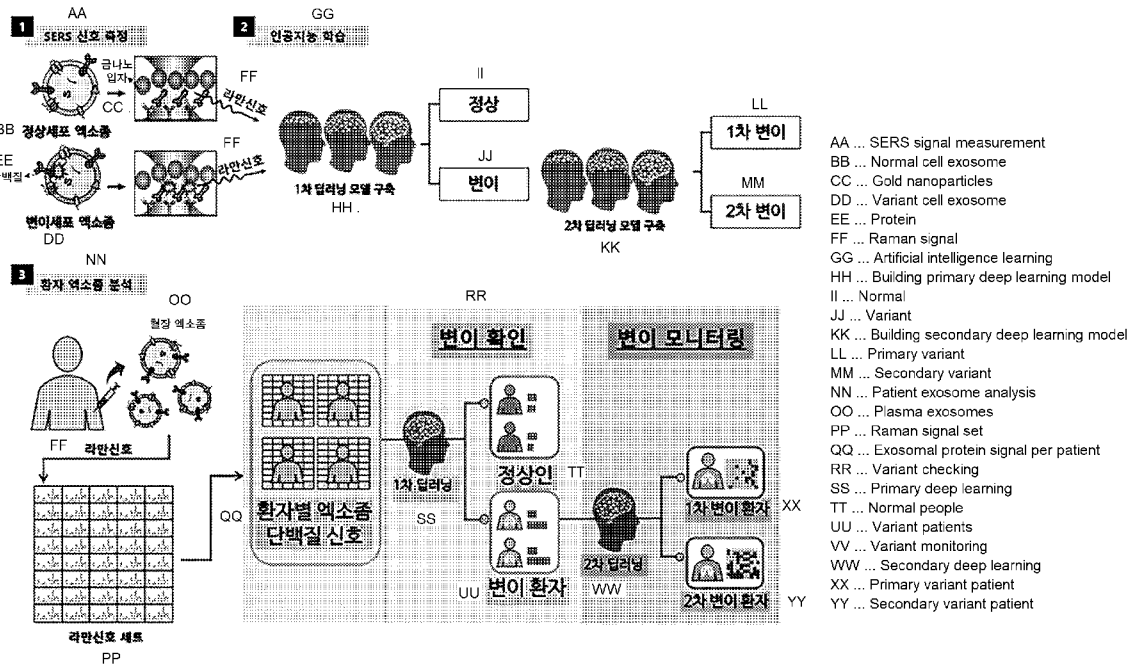
(10) 국제공개번호
WO 2024/158217 A1

- (51) 국제특허분류:
G16B 20/20 (2019.01) G06N 3/0464 (2023.01)
G16B 40/20 (2019.01) G16B 45/00 (2019.01)
- (21) 국제출원번호: PCT/KR2024/001171
- (22) 국제출원일: 2024년 1월 25일 (25.01.2024)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2023-0009725 2023년 1월 25일 (25.01.2023) KR
- (71) 출원인: 고려대학교 산학협력단 (KOREA UNIVERSITY RESEARCH AND BUSINESS FOUNDATION) [KR/KR]; 02841 서울특별시 성북구 안암로 145, Seoul (KR). 주식회사 엑스포트 (EXOPERT CORPORATION) [KR/KR]; 02580 서울특별시 동대문구 왕산로 15, 4층, Seoul (KR).

- (72) 발명자: 최연호 (CHOI, Yeonho); 06276 서울특별시 강남구 선릉로 221, 405동 603호, Seoul (KR). 김현구 (KIM, Hyun Koo); 06579 서울특별시 서초구 반포대로 39길 63-8, A동 402호, Seoul (KR). 김승민 (KIM, Seungmin); 02018 서울특별시 중랑구 공릉로13길 30, 102동 1704호, Seoul (KR). 최병현 (CHOI, Byeong Hyeon); 08529 서울특별시 금천구 남부순환로 112길 1-12, 803호, Seoul (KR). 신현구 (SHIN, Hyunku); 02853 서울특별시 성북구 고려대로 30, 808호, Seoul (KR).
- (74) 대리인: 특허법인 에스알비 (SRB IP & LAW FIRM); 06130 서울특별시 강남구 강남대로94길 53, 4층, 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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD,

(54) Title: METHOD AND APPARATUS FOR ANALYZING EXOSOMAL PROTEIN VARIANTS

(54) 발명의 명칭: 엑소솜 단백질의 변이 분석 방법 및 장치



(57) Abstract: The present invention relates to a method and a device that can classify surface-enhanced Raman scattering (SERS) signals of protein contained in an exosome, which are abundant in the blood and stably transport a biomarker, through a convolutional neural network (CNN)-based artificial intelligence algorithm, and detect and monitor exosomal protein variants by mathematically interpreting the derived classification results. Through the present invention, it is possible to overcome the limitations of signal heterogeneity in existing technologies, thereby enabling simple and rapid non-invasive liquid biopsy of an exosomal protein in plasma.

MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의
역내 권리의 보호를 위하여): ARIPO (BW, CV, GH, GM,
KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ,
UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ,
TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC,
ME, 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 발명은 혈액 내 풍부하게 존재하며 바이오마커를 안정적으로 운반하는 엑소솜에 함유된 단백질의
표면증강 라만 산란 (SERS) 신호를 합성곱 신경망 (CNN) 기반의 인공지능 알고리즘을 통해 분류하고, 도출된 분류
결과를 수학적으로 해석하여 엑소솜 단백질의 변이를 검출 및 모니터링할 수 있는 방법 및 장치에 관한 것으로,
본 발명을 통해 기존의 기술들이 가지고 있는 신호 이질성의 한계를 극복할 수 있으며, 이에 따라 혈장 내 엑소솜
단백질에 대한 간편하고 신속한 비침습적인 액체 생검이 가능하다.

명세서

발명의 명칭: 엑소좀 단백질의 변이 분석 방법 및 장치

기술분야

- [1] 본 특허출원은 2023년 01월 25일에 대한민국 특허청에 제출된 대한민국 특허출원 제10-2023-0009725호에 대하여 우선권을 주장하며, 상기 특허출원의 개시 사항은 본 명세서에 참조로서 삽입된다.
- [2] 본 발명은 과학기술정보통신부 지원 하에서 과제고유번호 1711159574, 과제번호 2021M3H4A4079630에 의해 이루어진 것으로서, 상기 과제의 연구관리전문기관은 한국연구재단, 연구사업명은 "나노소재기술개발", 연구과제명은 "나노 입자 프린팅 기술을 이용한 고감도 다중 바이오마커 검출 및 기계학습을 통한 pattern 분석 기반의 폐암 진단 기술개발", 주관기관은 고려대학교, 연구기간은 2022.02.01 ~ 2023.01.31이다.
- [3] 또한, 본 발명은 보건복지부 지원 하에서 과제고유번호 1465036580, 과제번호 HR14C0007060022에 의해 이루어진 것으로서, 상기 과제의 연구관리전문기관은 한국보건산업진흥원, 연구사업명은 "연구중심병원 육성 R&D", 연구과제명은 "엑소좀 기반 실시간 혈액암 조기진단 바이오칩 개발", 주관기관은 고려대학교 안암병원, 연구기간은 2022.01.01 ~ 2023.01.10이다.
- [4] 본 발명은 인공지능을 통해 엑소좀 단백질의 분광학 신호를 분석하여 변이 진단 및 추적 관찰하는 액체 생검 기법에 관한 것으로, 더욱 상세하게는, 본 발명은 혈액 내 풍부하게 존재하며 바이오마커를 안정적으로 운반하는 엑소좀에 함유된 단백질의 표면증강 라만 산란 (SERS) 신호를 합성곱 신경망 (CNN) 기반의 인공지능 알고리즘을 통해 분류하고, 도출된 분류 결과를 수학적으로 해석하여 엑소좀 단백질의 변이를 검출 및 모니터링할 수 있는 방법 및 장치에 관한 것이다.

[5]

배경기술

- [6] 엑소좀은 혈액, 소변, 침 등 체액 내 풍부하게 존재하는 미세 소포체 (extracellular vesicles)의 일종으로서 모든 세포에 의해 분비되는 것으로, 세포 간 상호작용, 압의 전이 등 다양한 생물학적 기작에 관여한다. 엑소좀은 30 ~ 200 nm 크기로 세포의 내포 작용으로 인해 내부에 내강 소낭이 생성되었다가, 다중 소포 엔도솜이 세포 표면과 결합해서 내강 소낭이 세포 밖으로 분비되는 외포 작용에 의해 생성된 것이기에 단백질과 miRNA 등 모세포의 특성을 대표할 수 있는 물질을 포함하며 유체 속에서도 안정성을 유지할 수 있다. 이러한 특성으로 인해 진단 의학 분야에서 액체 생검을 위한 바이오마커로서 엑소좀을 응용하고자 하는 노력이 지속되고 있지만, 거의 모든 생체유래물질의 고질적인 특징인 높은 이질성으로 인해 실질적인 응용에 어려움을 겪고 있다.

- [7] 한편, 표면증강 라만 산란 (SERS)은 플라즈모닉 나노구조체 사이의 나노갭에서 나타나는 강한 전자기장을 기반으로 하여 분자의 라만 신호가 10^7 - 10^8 배 이상 증폭되는 현상으로, 이를 활용하는 표면증강 라만 분광학을 통해 단백질의 라만 신호를 초고감도로 획득할 수 있어 돌연변이와 같은 국소적으로 일어나는 미세한 구조적 차이를 효과적으로 검출할 수 있다. 특히, SERS는 적은 농도의 시료에서도 높은 신호감도를 가지고 있기 때문에 비침습적인 액체생검을 통한 변이 진단 기술에 접목하기에 탁월한 적절성을 갖고 있다. 다만, 엑소솜 단백질의 SERS 신호는 단백질의 복잡한 구조, 신호 검출 프로브인 플라즈모닉 나노입자와의 상호작용, 및 부착면의 항체 분포 등의 이유로 인해 복잡하고 이질적인 신호가 검출된다는 한계점을 내포하고 있다.
- [8] 이러한 한계점의 해결을 위해 주성분분석법 (PCA), 부분최소제곱법 (PLS-DA) 등의 방법이 응용되었지만 환자 체액 내 존재하는 엑소솜은 야생형 세포에서 유래된 엑소솜과 변이 세포에서 유래된 엑소솜이 혼재되어 있는 상태이므로 PCA 또는 PLS-DA 기법 등을 통한 분석이 용이하지 않다. 또한, 변이 세포 유래 엑소솜은 상대적으로 체내에서 존재하는 농도가 굉장히 낮기 때문에 높은 농도의 야생형 단백질과 낮은 농도의 변이 단백질이 혼재된 분석체를 진단하는 기술의 개발이 지체되고 있는 실정이다.
- [9] 이에, 낮은 농도의 변이 엑소솜 단백질이 혼재된 검체로부터 획득되는 SERS 신호의 이질성을 극복할 수 있는 비침습적 생검 기법의 개발이 요구된다.

[10]

발명의 상세한 설명 기술적 과제

- [11] 이에, 본 발명자들은 인공지능 모델을 구축하여 야생형 세포 및 변이 세포가 혼합된 세포 배양액에서 유래된 엑소솜 단백질의 SERS 신호를 분석함으로써 효과적으로 검체 내의 변이 단백질 존재 여부를 파악하고, 변이 단백질의 변이 차수를 구분할 수 있음을 확인하였다.
- [12] 이에, 본 발명의 목적은 검체 내 엑소솜 단백질의 변이를 분석하는 방법을 제공하는 것이다.
- [13] 본 발명의 다른 목적은 검체 내 엑소솜 단백질의 변이를 분석하기 위한 컴퓨터 프로그램을 제공하는 것이다.
- [14] 본 발명의 또 다른 목적은 검체 내 엑소솜 단백질의 변이를 분석하기 위한 컴퓨팅 장치를 제공하는 것이다.

[15]

과제 해결 수단

- [16] 본 발명은 인공지능을 통해 엑소솜 단백질의 SERS 신호를 분석하여 변이 단백질을 진단 및 추적 관찰하는 방법 및 장치에 관한 것으로, 본 발명의 일 실시예에

따른 방법을 통해 신속하고 정확하게 검체 내 변이 단백질의 존재 여부를 파악하고 변이 단백질의 차수를 구분할 수 있다.

[17] 이하 본 발명을 더욱 자세히 설명하고자 한다.

[18]

[19] 본 발명의 일 양태는, 컴퓨팅 장치에 의해 수행되는, 검체 내 엑소솜 단백질의 변이를 분석하는 방법에 있어서, 검체 내 엑소솜 단백질의 표면증강 라만 산란 (SERS) 신호 데이터를 기초로 입력데이터를 생성하는 생성 단계; 및 미리 학습된 검출 모델을 이용하여, 입력데이터를 기반으로 엑소솜 단백질의 변이 여부를 판단하는 검출 단계;를 포함하는, 검체 내 엑소솜 단백질의 변이를 분석하는 방법에 관한 것이다.

[20] 본 명세서 상의 용어 "검체"는, 분석하고자 하는 엑소솜 단백질을 함유하고 있는 시료를 의미하는 것으로, 피검 대상으로부터 획득한 혈액, 혈장, 혈청, 림프액과 같은 관내 유체를 비롯하여 소변, 고름, 가래, 땀, 타액, 눈물, 콧물과 같은 점막 배출물 또는 샘 분비물과 같은 액체, 더 나아가 대변, 병변, 조직 생검 물질과 같은 고체 물질과 임의의 액체의 혼합물을 포함하는 것일 수 있으나, 이에 한정되는 것은 아니다.

[21] 본 명세서 상의 용어 "피검 대상"은, 본 발명의 일 실시예에 따른 분석 방법을 통해 엑소솜 단백질 변이 분석이 이루어지는 대상을 의미하는 것으로, 엑소솜 분비가 이루어지는 세포, 세포 집단, 조직, 기관 또는 생물일 수 있고, 구체적으로는 인간을 포함하는 영장류와 같은 포유동물일 수 있으나, 이에 한정되는 것은 아니다.

[22] 본 명세서 상의 용어 "엑소솜 단백질"은, 검체 내의 엑소솜에 포함된 단백질을 의미하는 것으로, 엑소솜의 인지질 막에 연결된 막 단백질 또는 엑소솜의 내강에 함유된 단백질을 총칭하거나, 또는 엑소솜에 포함된 총 단백질 중 일련의 선별과정을 통해 선별된 한 개 또는 복수 개의 표적 단백질을 의미하는 것일 수 있다.

[23] 본 명세서 상의 용어 "변이"는, 한 종의 개체에서 다수를 차지하는 형질인 야생형과 달리 발현되는 형질로서, 그로 인해 변이 개체의 표현 형질 또는 행동 양식이 야생형 개체에 비해 유의미한 차이가 발생하게 되는 것을 의미한다. 여기서, 개체는 피검 대상과 마찬가지로 엑소솜 분비가 이루어지는 세포, 세포 집단, 조직, 기관 또는 생물일 수 있고, 구체적으로는 인간을 포함하는 영장류와 같은 포유동물일 수 있다. 여기서, 개체에서 발현되는 형질은 유전자에 따라 발현되는 단백질을 포함할 수 있다.

[24] 본 명세서 상의 용어 "입력데이터"는, 본 발명의 일 실시예에 따른 신경망 모델에 가장 먼저 입력되는 데이터를 의미하는 것으로, 입력데이터는 검체 내 엑소솜 단백질로부터 측정된 SERS 신호가 전자적으로 변환된 SERS 신호 데이터에 대해 일련의 연산이 수행됨으로써 생성될 수 있다. 입력데이터의 형태는 특별히 한정되지 않으며, 예를 들어, 텐서 데이터일 수 있다.

- [25] 본 명세서 상의 용어 "텐서 데이터"는, 하나 이상의 값의 다차원 배열로 이루어진 데이터 형태를 의미한다. 본 발명의 일 실시예에 따른 신경망 모델에 입력되는 입력데이터의 형태는 0차원, 1차원, 2차원 또는 그 이상의 차원의 텐서 데이터일 수 있고, 예를 들어, 1차원 벡터 데이터일 수 있으나, 이에 한정되는 것은 아니다.
- [26] 본 명세서 상의 용어 "학습"은, 신경망이 도출해야 하는 결과 또는 판단에 대해, 신경망을 구성하는 알고리즘 하에 신경망이 최적의 결과값을 도출하거나 판단을 내릴 수 있도록, 알고리즘에 따른 각 동작을 최적화하는 과정을 의미한다. 본 발명의 일 실시예에 따른 신경망 모델의 학습은 학습 데이터에 정답이 라벨링된 학습 데이터를 이용하는 지도 학습 (supervised learning), 학습 데이터에 정답이 라벨링되지 않은 비지도 학습 (unsupervised learning), 정답이 라벨링된 학습 데이터와 라벨링되지 않은 학습 데이터를 모두 이용하는 반지도 학습 (semi-supervised learning), 또는 보상을 통해 모델의 상태별 행동이 최적화되는 강화학습 (reinforcement learning) 중 적어도 하나의 방식으로 이루어질 수 있으며, 예를 들어, 지도 학습을 통해 신경망 모델의 학습이 이루어질 수 있다.
- [27] 본 발명의 일 구현예에서, 검체 내 엑소좀 단백질의 변이를 분석하는 방법은, 검체 내 엑소좀 단백질로부터 SERS 신호를 측정하는 신호 측정 단계;를 포함하는 것일 수 있다.
- [28] 본 발명의 일 구현예에서, 생성 단계는, 복수의 SERS 신호 데이터를 통계적 처리를 통해 평균 신호로 변환하는 통계적 처리 단계;를 포함하는 것일 수 있다.
- [29] 본 명세서 상의 용어 "평균 신호"는, 검체 내 엑소좀 단백질로부터 복수 회의 SERS 신호 측정을 수행한 후, 복수의 SERS 신호 또는 복수의 SERS 신호 데이터를 통계적으로 처리하여 도출되는 신호를 의미하는 것으로, 통계적 처리에는 신호의 평균 데이터 획득 또는 신호의 표준편차 데이터 획득이 포함될 수 있다. 이때, 복수 회의 SERS 신호 측정은 검체의 전체적인 특징을 추출하기에 충분한 회수로 수행될 수 있으며, 예를 들어, 80 내지 100회의 SERS 신호 측정을 통해 평균 신호를 도출할 수 있으나, 이에 한정되는 것은 아니다.
- [30] 본 발명의 일 구현예에서, 통계적 처리 단계는, 80개 내지 100개의 SERS 신호 데이터를 통계적 처리를 통해 평균 신호로 변환하는 것일 수 있다.
- [31] 본 발명의 일 구현예에서, 검출 단계는, 미리 학습된 검출 모델을 이용하여, 엑소좀 단백질의 변이 여부에 대한 출력정보를 산출하는 산출 단계; 및 출력정보를 기초로, 검체 내 엑소좀 단백질과 야생형 엑소좀 단백질 및 변이 엑소좀 단백질의 상대적 유사도를 결정하는 유사도 결정 단계;를 포함하는 것일 수 있다.
- [32] 본 명세서 상의 용어 "출력정보"는, 본 발명의 일 실시예에 따른 신경망 모델에 입력된 입력데이터에 대해 수행되는 일련의 연산에 따라 출력되는 데이터를 의미한다. 출력정보의 형태는 입력데이터와 동일하거나 상이할 수 있으며, 예를 들어, 출력정보의 형태는 0차원 스칼라 데이터 또는 1차원 벡터 데이터일 수 있으나, 이에 한정되는 것은 아니다.

- [33] 본 명세서 상의 용어 "상대적 유사도"는, 검체 내 엑소솜 단백질이 야생형 엑소솜 단백질 및 변이 엑소솜 단백질 중 어느 것과 더 유사한지 판단하기 위한 척도를 의미하는 것으로, 검체 내 엑소솜 단백질에 관한 데이터를 야생형 및 변이 단백질에 관한 데이터와 각각 비교하여 개별 유사도를 도출하고, 도출된 개별 유사도를 서로 비교하는 것으로 상대적 유사도를 결정할 수 있다.
- [34] 이때, 개별 유사도의 도출 시 데이터 간의 유사도를 정량화하기 위한 임의의 기법을 이용할 수 있으며, 예를 들어, 검체 내 엑소솜 단백질에 관한 데이터와 야생형 또는 변이 단백질에 관한 데이터 간의 마할라노비스 거리 (Mahalanobis distance) 계산을 통해 개별 유사도가 도출될 수 있다.
- [35] 본 발명의 일 구현예에서, 출력정보는 길이가 128인 1차원의 텐서 데이터인 것일 수 있다.
- [36] 본 발명의 일 구현예에서, 검출 모델은, 야생형 엑소솜 단백질로부터 생성된 SERS 신호 데이터 및 변이 엑소솜 단백질로부터 생성된 SERS 신호 데이터를 포함하는 학습 데이터, 및 학습 데이터와 대응되는 엑소솜 단백질 변이 여부에 관한 검출 라벨에 기초하여 지도 학습되며, 지도 학습은 검출 모델을 이용하여 학습 데이터로부터 생성한 학습 정보, 및 검출 라벨에 기초하여 수행되는 것일 수 있다.
- [37] 본 명세서 상의 용어 "검출 라벨"은, 본 발명의 일 실시예에 따른 검출 모델의 학습을 위해, 야생형 엑소솜 단백질로부터 생성된 SERS 신호 데이터 및 변이 엑소솜 단백질로부터 생성된 SERS 신호 데이터를 포함하는 학습 데이터에 정답으로서 라벨링되는 이진 데이터를 의미한다. 예를 들어, 본 발명의 일 실시예에 따른 검출 모델은 야생형 단백질에 대해서는 1, 변이 단백질에 대해서는 0의 값이 할당된 검출 라벨 및 학습 데이터를 통해 지도 학습될 수 있다.
- [38] 본 발명의 일 구현예에서, 검체 내 엑소솜 단백질의 변이를 분석하는 방법은, 검출 단계에서 엑소솜 단백질이 변이된 것으로 판단된 경우, 미리 학습된 구분 모델을 이용하여, 입력데이터를 기반으로 변이 엑소솜 단백질의 변이 차수를 구분하는 구분 단계;를 포함하는 것일 수 있다.
- [39] 본 명세서 상의 용어 "변이 차수"는, 변이가 발생한 개체에 있어 몇 차례의 변이가 발생한 것인지 구분하는 것을 의미하며, 예를 들어, 개체에 최초로 발생한 변이는 1차 변이, 그 이후로 발생한 변이에 대해서는 2차 이상의 변이로 변이 차수를 구분할 수 있다. 2차 이상의 변이를 판단할 때, 1차 변이가 발생한 단백질과 동일한 단백질의 다른 아미노산 서열에 발생한 후속 변이 및 1차 변이가 발생한 단백질과 상이한 단백질에 발생한 후속 변이를 모두 2차 이상의 변이로 지칭할 수 있다.
- [40] 본 발명의 일 구현예에서, 구분 단계는, 검출 단계에서 엑소솜 단백질이 변이된 것으로 판단된 경우, 검출 단계 종결 이후 소정의 기간이 경과한 후, 동일한 피검 대상으로부터 새로 획득된 검체의 엑소솜 단백질에 대해 수행되는 것일 수 있다.

- [41] 본 발명의 일 구현예에서, 구분 단계는, 전체 SERS 신호 데이터 중 2차 이상의 변이가 존재하는 것으로 판단된 SERS 신호 데이터의 비율을 계산하여 변이정보를 생성하는 변이정보 생성 단계;를 포함하는 것일 수 있다.
- [42] 본 명세서 상의 용어 "전체 SERS 신호 데이터"는, 검체 내 엑소좀 단백질로부터 측정된 복수의 SERS 신호를 바탕으로 생성된 복수의 SERS 신호 데이터의 집합을 의미하며, 평균 신호를 도출하기 위해 검체 내 엑소좀 단백질로부터 생성된 복수의 SERS 신호 데이터의 경우도 전체 SERS 신호 데이터에 해당하는 것으로 볼 수 있다. 본 명세서 상에서 '전체 SERS 신호 데이터'는 'SERS 신호 데이터 세트'와 상호 교환 가능하게 사용될 수 있다.
- [43] 본 명세서 상의 용어 "변이정보"는, 엑소좀 단백질이 변이된 것으로 판단된 피검 대상의 검체 내 엑소좀 단백질로부터 획득한 전체 SERS 신호 데이터 중 2차 이상의 변이가 존재하는 것으로 판단된 SERS 신호 데이터의 비율을 의미한다.
- [44] 이때, 변이정보의 비율에 대해 특정 컷오프 (cut-off) 값을 적용하여, 변이정보의 비율이 해당 컷오프 값 이상인 경우 피검 대상에서 2차 이상의 변이가 발생한 것으로 구분할 수 있다.
- [45] 본 발명의 일 구현예에서, 전체 SERS 신호 데이터는 검체 내 엑소좀 단백질로부터 측정된 100개의 SERS 신호를 기반으로 생성되는 것일 수 있다.
- [46] 본 발명의 일 구현예에서, 구분 모델은, 엑소좀 단백질의 변이 차수에 따른 SERS 신호 데이터를 포함하는 차수 학습 데이터, 및 차수 학습 데이터와 대응되는 변이 차수에 관한 구분 라벨에 기초하여 지도 학습되며, 지도 학습은 구분 모델을 이용하여 차수 학습 데이터로부터 생성한 학습 정보, 및 구분 라벨에 기초하여 수행되는 것일 수 있다.
- [47] 본 명세서 상의 용어 "구분 라벨"은, 본 발명의 일 실시예에 따른 구분 모델의 학습을 위해, 1차 변이가 발생한 엑소좀 단백질로부터 생성한 SERS 신호 데이터 및 2차 이상의 변이가 발생한 엑소좀 단백질로부터 생성한 SERS 신호 데이터를 포함하는 차수 학습 데이터에 정답으로서 라벨링되는 스칼라 데이터를 의미한다.
- [48] 구분 라벨의 개수는 구분 모델의 학습을 최적화하기 위한 임의의 개수로 설정될 수 있으며, 예를 들어, 1차 변이에 대해서는 0, 2차 변이에 대해서는 1을 할당할 수 있다.
- [49] 1차 변이와 2차 이상의 변이를 동시에 학습하기 위해 세 개 이상의 구분 라벨을 설정하는 경우 신경망 모델이 해결해야 할 문제는 다중 분류 문제가 되는데, 검체 내 3차 이상의 변이 단백질이 야생형 단백질 또는 1차 및 2차 변이 단백질 농도에 비해 훨씬 낮은 농도로 존재한다는 점으로 인해 신경망 모델의 다중 분류 학습 효율 및 성능이 저하될 수 있다.
- [50] 따라서, 변이 구분 모델의 학습 효율 및 성능을 고려할 때, 3차 이상의 변이를 구분하는 경우에도 다중 라벨이 아닌 이진 라벨을 단계적으로 할당함으로써 고차원의 변이를 효과적으로 구분할 수 있다.

- [51] 본 발명의 다른 양태는, 저장 매체에 저장된 컴퓨터 프로그램으로서, 컴퓨터 프로그램은 하나 이상의 프로세서에서 실행되는 경우, 검체 내 엑소좀 단백질의 변이를 분석하기 위한 동작들을 수행하며, 동작들은 하기를 포함하는 것인, 저장 매체에 저장된 컴퓨터 프로그램에 관한 것이다: 검체 내 엑소좀 단백질의 SERS 신호 데이터를 기초로 입력데이터를 생성하는 동작; 및 미리 학습된 검출 모델을 이용하여, 상기 입력데이터를 기반으로 엑소좀 단백질의 변이 여부를 판단하는 동작.
- [52] 본 발명의 또 다른 양태는, 검체 내 엑소좀 단백질의 변이를 분석하기 위한 컴퓨팅 장치로서, 하나 이상의 코어를 포함하는 프로세서; 및 메모리;를 포함하고, 프로세서는, 검체 내 엑소좀 단백질의 SERS 신호 데이터를 기초로 입력데이터를 생성하고, 및 미리 학습된 검출 모델을 이용하여, 입력데이터를 기반으로 엑소좀 단백질의 변이 여부를 판단하는 것인, 컴퓨팅 장치에 관한 것이다.

[53]

발명의 효과

- [54] 본 발명에 따른 엑소좀 단백질의 변이 분석 방법 및 장치를 통해 SERS 신호 분석에 있어서 기존의 기술들이 가지고 있는 신호 이질성의 한계를 극복할 수 있으며, 이에 따라 혈장 내 엑소좀 단백질에 대한 간편하고 신속한 비침습적인 액체 생검이 가능하므로, 본 발명은 질병 진단을 위한 종래의 침습적·비침습적 생검 기술을 대체할 수 있을 것으로 기대된다.

[55]

도면의 간단한 설명

- [56] 도 1은 본 발명의 일 실시예에 따른 검체 내 엑소좀 단백질의 변이를 분석하기 위한 동작들을 수행하는 컴퓨팅 장치의 블록구성도이다.
- [57] 도 2는 본 발명의 일 실시예에 따른 검출 모델 또는 구분 모델을 구성하는 신경망 모델의 구조를 나타낸 개략도이다.
- [58] 도 3은 본 발명의 일 실시예에 따라 검체 내 엑소좀 단백질의 변이를 분석하는 과정을 설명하기 위한 흐름도이다.
- [59] 도 4는 본 발명의 일 실시예에 따라 검체 내 엑소좀 단백질의 변이를 분석하는 과정을 나타낸 모식도이다.
- [60] 도 5는 야생형 엑소좀 단백질 및 폐암세포 유래 변이 엑소좀 단백질에서 획득한 SERS 신호 추이를 나타낸 그래프이다.
- [61] 도 6은 야생형 세포 및 변이 세포 유래 엑소좀 단백질 부착 기관으로부터 SERS 신호 데이터 세트를 확보하는 과정을 나타낸 모식도이다.
- [62] 도 7는 본 발명의 일 실시예에 따른 검출 모델의 학습 정확도 및 손실값을 나타낸 그래프이다.
- [63] 도 8는 본 발명의 일 실시예에 따른 구분 모델의 학습 정확도 및 손실값을 나타낸 그래프이다.

- [64] 도 9는 본 발명의 일 실시예에 따른 검출 모델을 통해 검체 내 엑소솜 단백질과 야생형 엑소솜 단백질 및 변이 엑소솜 단백질 간의 상대적 유사도를 결정하여 변이 여부를 검출하는 과정을 나타낸 모식도이다.
- [65] 도 10는 본 발명의 일 실시예에 따른 검출 모델을 통해 결정된 엑소솜 단백질 SERS 신호의 상대적 유사도를 바탕으로 정상인과 폐암 환자를 분류한 결과를 나타낸 그래프이다.
- [66] 도 11는 본 발명의 일 실시예에 따른 검출 모델의 변이 엑소솜 단백질 검출 성능을 나타낸 ROC 그래프이다.
- [67] 도 12는 환자의 장기 치료에 따라 발생하는 변이 차수 변화를 본 발명의 일 실시예에 따른 구분 모델을 통해 구분하는 과정을 나타낸 모식도이다.
- [68] 도 13는 본 발명의 일 실시예에 따른 구분 모델을 통해 생성한 변이정보를 바탕으로 폐암 환자를 1차 변이 환자 및 2차 변이 환자로 구분한 결과를 나타낸 그래프이다.
- [69] 도 14는 본 발명의 일 실시예에 따른 구분 모델을 통해 2차 변이 환자로 구분된 피검 대상에 있어, 1차 변이 진단 시점 (t_1) 및 2차 변이 진단 시점 (t_2)에 각각 획득한 100개의 엑소솜 단백질 SERS 신호 중 2차 변이에 해당하는 것으로 판단된 신호의 비율을 나타낸 도면이다.
- [70] 도 15는 본 발명의 일 실시예에 따른 구분 모델을 통해 2차 변이 환자로 구분된 피검 대상에 있어, 1차 변이 진단 시점 (t_1) 대비 2차 변이 진단 시점 (t_2)에서의 2차 변이 신호 증감 비율을 나타낸 그래프이다.

[71]

발명의 실시를 위한 최선의 형태

- [72] 본 발명은 컴퓨팅 장치에 의해 수행되는, 검체 내 엑소솜 단백질의 변이를 분석하는 방법에 있어서,
- [73] 검체 내 엑소솜 단백질의 표면증강 라만 산란 (SERS) 신호 데이터를 기초로 입력데이터를 생성하는 생성 단계; 및
- [74] 미리 학습된 검출 모델을 이용하여, 상기 입력데이터를 기반으로 엑소솜 단백질의 변이 여부를 판단하는 검출 단계;
- [75] 를 포함하는, 검체 내 엑소솜 단백질의 변이를 분석하는 방법에 관한 것이다.

[76]

발명의 실시를 위한 형태

- [77] 이하, 본 발명에 따른 바람직한 실시 형태를 첨부된 도면을 참조하여 상세하게 설명한다. 첨부된 도면과 함께 이하에 개시될 상세한 설명은 본 발명의 예시적인 실시 형태를 설명하고자 하는 것으로, 본 발명이 실시될 수 있는 유일한 실시 형태를 나타내고자 하는 것이 아니고, 본 발명의 요지에 따라 본 발명의 범위가 이들 실시예에 의해 제한되지 않는다.

- [78] 몇몇의 경우, 본 발명의 개념이 모호해지는 것을 피하기 위하여 공지의 구조 및 장치는 생략되거나, 각 구조 및 장치의 핵심기능을 중심으로 한 블록도 형식으로 도시될 수 있다. 또한, 본 명세서 전체에서 동일한 구성요소에 대해서는 동일한 도면 부호를 사용하여 설명한다.
- [79] 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.
- [80] 또한, 명세서 전체에서, 어떤 구성이 다른 구성과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, "그 중간에 다른 구성을 사이에 두고 연결"되어 있는 경우도 포함한다.
- [81] 용어 "또는"은 배타적 "또는"이 아니라 내포적 "또는"을 의미하는 것으로 의도된다. 즉, 구성 간의 결합 또는 이용이 달리 특정되지 않거나 문맥상 명확하지 않은 경우에, 즉, X가 A를 이용하거나; X가 B를 이용하거나; 또는 X가 A 및 B 모두를 이용하는 경우, "X는 A 또는 B를 이용한다"가 이들 경우들 어느 것으로도 적용될 수 있고, X가 A와 연결되거나; X가 B와 연결되거나; 또는 X가 A 및 B 모두와 연결되는 경우, "X는 A 또는 B와 연결된다"가 이들 경우들 어느 것으로도 적용될 수 있다.
- [82] 또한, 명세서에 기재된 "...부"의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다. 나아가, "일", "하나" 및 유사 관련어는 본 발명을 기술하는 문맥에 있어서 본 명세서에 달리 지시되거나 문맥에 의해 분명하게 반박되지 않는 한, 단수 및 복수 모두를 포함하는 의미로 사용될 수 있다.
- [83] 본 명세서 상에서 실시예들과 관련되어 설명된 다양한 예시적 논리적 블록들, 구성들, 모듈들, 회로들, 수단들, 로직들, 및 알고리즘 단계들이 전자 하드웨어, 컴퓨터 소프트웨어, 또는 양쪽 모두의 조합들로 구현될 수 있음을 인식해야 한다. 하드웨어 및 소프트웨어의 상호교환성을 명백하게 예시하기 위해, 다양한 예시적 컴포넌트들, 블록들, 구성들, 수단들, 로직들, 모듈들, 회로들, 및 단계들은 그들의 기능성 측면에서 일반적으로 위에서 설명되었다. 그러한 기능성이 하드웨어로 또는 소프트웨어로서 구현되는지 여부는 전반적인 시스템에 부과된 특정 어플리케이션(application) 및 설계 제한들에 달려 있다. 숙련된 기술자들은 각각의 특정 어플리케이션들을 위해 다양한 방법들로 설명된 기능성을 구현할 수 있다. 다만, 그러한 구현의 결정들이 본 개시내용의 영역을 벗어나게 하는 것으로 해석되어서는 안 된다.
- [84]
- [85] 도 1은 본 발명의 일 실시예에 따른 검체 내 엑소좀 단백질의 변이를 분석하기 위한 동작들을 수행하는 컴퓨팅 장치의 블록구성도이다.

- [86] 도 1을 참조하면, 본 발명의 일 실시예에 따른 검체 내 엑소솜 단백질의 변이를 분석하기 위한 동작들을 수행하는 컴퓨팅 장치 (100)는 프로세서 (101) 및 메모리 (102)를 포함할 수 있다.
- [87] 프로세서 (101)는, 검체 내 엑소솜 단백질의 표면증강 라만 산란 (SERS) 신호 데이터를 기초로 입력데이터를 생성하는 동작을 수행할 수 있다.
- [88] SERS 신호 데이터는 피검 대상에서 획득한 검체 내 엑소솜 단백질의 SERS 신호에 대해 일련의 전처리를 거쳐 생성될 수 있고, 예를 들어, 표준화 (standardization) 또는 최소-최대 정규화 (min-max normalization)을 통해 SERS 신호가 신경망 모델 학습에 적합한 데이터로 변환될 수 있다.
- [89] 프로세서 (101)는, SERS 신호 데이터를 텐서 형태의 입력데이터로 변환할 수 있다. 입력데이터는 임의의 데이터 형태를 가질 수 있으며, 예를 들어, 1차원의 벡터 형태 또는 2차원의 행렬 형태일 수 있다.
- [90] 프로세서 (101)는, 복수의 SERS 신호 데이터를 통계적 처리를 통해 평균 신호로 변환하는 동작을 수행할 수 있다. 복수의 SERS 신호 데이터는 하나의 검체에 대해 복수 회 측정된 SERS 신호를 전처리함으로써 생성될 수 있고, 하나의 검체로부터 생성된 복수의 SERS 신호 데이터는 SERS 신호 데이터 세트를 구성할 수 있다. 복수의 SERS 신호 데이터를 평균 신호로 변환하기 위한 통계적 처리 기법은 특정 기법에 제한되지 않으며, 예를 들어, 복수의 SERS 신호 데이터의 각 성분의 평균 및 표준편차를 도출하여 평균 신호로 변환할 수 있다.
- [91] 프로세서 (101)는, 미리 학습된 검출 모델을 이용하여, 입력데이터를 기반으로 엑소솜 단백질의 변이 여부를 판단하는 동작을 수행할 수 있다. 미리 학습된 검출 모델은 신경망 모델일 수 있고, 특히 심층 신경망 모델일 수 있다.
- [92] 심층 신경망은 입력 계층 (input layer)과 출력 계층 (output layer) 사이에 여러 개의 은닉 계층 (hidden layer)이 존재하며, 각 계층을 구성하는 하나 이상의 노드 (node)와 노드를 연결하는 하나 이상의 에지 (edge)로 구성된 인공 신경망으로, 입력층에서 출력층에 이르는 연산 과정을 통해 복잡한 비선형 관계를 모델링할 수 있다. 이러한 신경망은 '네트워크' 또는 '뉴럴 네트워크'로, 심층 신경망은 '딥 뉴럴 네트워크'로 각각 지칭되어 상호 교환 가능하게 사용될 수 있다.
- [93] 심층 신경망은 입력층에서 출력층에 이르는 연산 과정의 알고리즘에 따라 합성곱 신경망 (convolutional neural network; CNN), 순환 신경망 (recurrent neural network; RNN), 오토 인코더 (auto encoder), 생성적 적대 신경망 (Generative Adversarial Networks), 제한 볼츠만 머신 (restricted boltzmann machine; RBM), 심층 신뢰망 (deep belief network; DBN), Q 네트워크, U네트워크, 또는 삼 네트워크 등으로 분류될 수 있으나, 이에 한정되는 것은 아니다.
- [94] 검출 모델이 심층 신경망 모델인 경우, 검출 모델의 구조는 임의의 심층 신경망 모델의 조합으로 구성될 수 있으며, 예를 들어, 검출 모델은 다층의 합성곱 단위 블록을 포함하는 합성곱 신경망 및 다층의 완전연결 계층을 포함하는 완전연결 신경망 조합으로 구성될 수 있으나, 이에 한정되는 것은 아니다.

- [95] 합성곱 신경망은 동물의 시각 피질 조직에서 영감을 받아, 이미지, 소리 등과 같은 다차원 패턴을 갖는 데이터를 처리하기 위해 고안된 심층 신경망 모델의 일종으로, 합성곱 신경망은 일반적으로 합성곱 계층 (convolution layer), 풀링 계층 (pooling layer) 및 완전연결 계층 (fully-connected layer)을 포함할 수 있다. 합성곱 신경망 내에 하나 이상의 합성곱 계층 및 풀링 계층이 반복적으로 존재할 수 있으며, 합성곱 신경망에 입력된 입력값은 이러한 계층 구조를 거치며 출력값으로 변환될 수 있다. 대표적인 합성곱 신경망의 구조로서 AlexNet, VGGNet, ResNet과 같은 구조가 있으나, 본 발명의 일 실시예에 따른 합성곱 신경망의 구조는 이에 한정되는 것은 아니다.
- [96] 완전연결 신경망 (fully-connectec network)은 입력 계층, 출력 계층 및 그 사이에 존재하는 여러 개의 은닉 계층을 포함할 수 있다. 완전연결 신경망에서, 인접한 각 계층의 하나 이상의 노드는 인접한 계층 간에 입력 노드 및 출력 노드 관계를 형성할 때, 입력 노드로 구성된 계층의 모든 노드는 각각 에지를 통해 출력 노드로 구성된 계층의 모든 노드와 연결될 수 있다.
- [97] 이때, 노드는 하나의 데이터 단위를 나타내는 객체를 의미하는 것일 수 있으며, 하나의 에지를 통해 연결된 입력 노드 및 출력 노드의 관계에서, 출력 노드의 데이터는 입력 노드에 입력된 데이터에 기초하여 그 값이 결정될 수 있다. 노드는 '뉴런'으로 지칭되어 상호 교환 가능하게 사용될 수 있다.
- [98] 또한, 에지는 가중치를 가질 수 있으며, 추가적으로 바이어스를 가질 수 있다. 가중치 또는 바이어스는 가변적일 수 있으며, 에지는 입력 노드에 입력된 값에 대해 가중치 또는 바이어스를 반영하는 연산을 수행하여 출력 노드의 값을 결정할 수 있다. 에지는 '링크'로 지칭되어 상호 교환 가능하게 사용될 수 있다.
- [99] 프로세서 (101)는, 미리 학습된 검출 모델을 이용하여, 엑소좀 단백질의 변이 여부에 대한 출력정보를 산출하는 동작을 수행할 수 있다. 구체적으로, 검체 내 엑소좀 단백질의 SERS 신호로부터 생성된 입력데이터를 검출 모델에 입력하면 최종 출력으로서 검체 내 엑소좀 단백질의 변이 여부에 대한 출력정보가 산출될 수 있고, 이때 출력정보의 데이터 형태는 텐서로서 0차원, 1차원, 2차원 또는 3차원 이상의 형태를 지닐 수 있고, 예를 들어, 출력정보는 0차원 스칼라 데이터 또는 1차원 벡터 데이터의 형태일 수 있다.
- [100] 프로세서 (101)는, 출력정보를 기초로, 검체 내 엑소좀 단백질과 야생형 엑소좀 단백질 및 변이 엑소좀 단백질의 상대적 유사도를 결정하는 동작을 수행할 수 있다. 상대적 유사도를 결정하기 위한 출력정보 산출 시, 단일한 성분을 갖는 스칼라 또는 둘 이상의 성분을 갖는 벡터의 형태로 출력정보가 산출될 수 있다. 이때, 검체 내 엑소좀 단백질은 변이 단백질과 야생형 또는 정상 단백질이 혼재되어 있는 상태로서 단순한 이진 분류를 통해 변이 여부가 판단되기 어렵기 때문에, 변이 여부 판단을 위한 상대적 유사도 결정 시 출력정보는 벡터의 형태로 산출될 수 있다.

- [101] 상대적 유사도 결정을 위해 검체 내 엑소좀 단백질에 관한 데이터와 야생형 또는 변이 단백질에 관한 데이터 간의 마할라노비스 거리를 계산하는 경우, 각 데이터에 대해 벡터 형태로 출력되는 출력정보의 개별 성분들의 평균 및 공분산을 도출한 후, 각 출력정보의 평균 차이에 공분산을 반영함으로써 마할라노비스 거리를 계산할 수 있다.
- [102] 프로세서 (101)는, 검출 단계에서 엑소좀 단백질이 변이된 것으로 판단된 경우, 미리 학습된 구분 모델을 이용하여, 입력데이터를 기반으로 변이 엑소좀 단백질의 변이 차수를 구분하는 동작을 수행할 수 있다. 미리 학습된 구분 모델은 신경망 모델일 수 있고, 특히 심층 신경망 모델일 수 있다.
- [103] 구분 모델이 심층 신경망 모델인 경우, 구분 모델의 구조는 임의의 심층 신경망 모델의 조합으로 구성될 수 있으며, 예를 들어, 구분 모델은 다층의 합성곱 단위 블록을 포함하는 합성곱 신경망 및 다층의 완전연결 계층을 포함하는 완전연결 신경망 조합으로 구성될 수 있으나, 이에 한정되는 것은 아니다.
- [104] 프로세서 (101)는, 전체 SERS 신호 데이터 중 2차 이상의 변이가 존재하는 것으로 판단된 SERS 신호 데이터의 비율을 계산하여 변이정보를 생성하는 동작을 수행할 수 있다. 변이정보는 구분 모델에 입력된 입력데이터, 또는 별도로 생성된 입력데이터를 기초로 생성될 수 있다.
- [105] 프로세서 (101)는, 변이정보 생성을 위해 전체 SERS 신호 데이터 중 2차 이상의 변이가 존재하는 것으로 판단된 SERS 신호 데이터의 비율을 계산할 때, 입력데이터로부터 2차 이상의 변이를 판단하기 위한 출력데이터를 산출할 수 있다. 2차 이상의 변이를 판단하기 위한 출력데이터는 텐서의 형태로서 0차원 스칼라 데이터, 1차원 벡터 데이터, 2차원 행렬 데이터 또는 3차원 이상의 형태를 지닐 수 있고, 예를 들어, 특정 입력데이터에 대한 2차 이상의 변이 존재 여부를 이진 분류하기 위한 스칼라 값을 가질 수 있다.
- [106] 프로세서 (101)는, 신경망 모델의 학습을 위한 입력데이터의 생성, 오차 계산 및 역전파 (backpropagation)를 이용한 신경망 모델의 연산을 위한 가중치 최적화 등의 신경망의 학습을 위한 계산을 수행할 수 있다.
- [107] 프로세서 (101)는, 하나 이상의 코어로 구성될 수 있으며, 컴퓨팅 장치의 중앙 처리 장치 (central processing unit; CPU), 그래픽 처리 장치 (graphics processing unit; GPU), 텐서 처리 장치(tensor processing unit; TPU) 등의 데이터 분석, 딥러닝을 위한 프로세서를 포함할 수 있다.
- [108] 또한 프로세서 (101)는, 메모리 (102)에 저장된 컴퓨터 프로그램을 판독하여 본 발명의 일 실시예에 따른 신경망 모델의 학습을 위한 데이터 처리를 수행할 수 있고, 프로세서 (101)는, 신경망 모델의 학습을 위한 연산을 수행할 수 있다.
- [109] 메모리 (102)는 프로세서 (101)가 생성하거나 결정한 임의의 형태의 정보를 저장할 수 있는 것으로, 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(Random

Access Memory; RAM), SRAM(Static Random Access Memory), 롬(Read-Only Memory; ROM), EEPROM(Electrically Erasable Programmable Read-Only Memory), PROM(Programmable Read-Only Memory), 자기 메모리, 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다.

- [110] 본 발명의 일 실시예에 따른 검체 내 엑소좀 단백질의 변이를 분석하기 위한 동작들을 수행하는 컴퓨팅 장치 (100)는 필요에 따라 검체 내 엑소좀 단백질의 SERS 신호가 포함된 정보 및 기타 데이터를 송수신하기 위한 네트워크부, 및 컴퓨팅 과정에서 이용되는 데이터를 저장 및 관리하기 위한 데이터베이스를 더 포함할 수 있다.
- [111] 따라서, 본 발명의 일 실시예에 따른 검체 내 엑소좀 단백질의 변이를 분석하는 동작들을 수행하는 컴퓨팅 장치 (100)는 검체 내 엑소좀 단백질의 SERS 신호를 측정할 수 있는 타 컴퓨팅 장치, 단말 또는 전자 장비 등과 유·무선의 정보 교환이 가능할 수 있다. 그리고, 야생형 엑소좀 단백질의 SERS 신호 데이터 또는 변이 엑소좀 단백질의 SERS 신호 데이터가 저장된 타 컴퓨팅 장치, 단말, 저장용 메모리 또는 클라우드 (cloud) 서버 등과 유·무선의 정보 교환이 가능할 수 있다.
- [112] 또한, 엑소좀 단백질의 SERS 데이터 및 이로부터 생성한 입력데이터, 검출 모델 및 구분 모델의 가중치 데이터, 입력데이터로부터 산출되는 출력 정보, 검체 내 엑소좀 단백질과 야생형 단백질 및 변이 단백질 간의 상대적 유사도 및 검체 내 엑소좀 단백질의 변이 정보 등의, 검체 내 엑소좀 단백질의 변이를 분석하기 위한 동작을 수행하는 과정에서 생성되는 데이터는 본 발명의 일 실시예에 따른 컴퓨팅 장치 (100)의 데이터베이스에 저장될 수 있으며, 데이터베이스에 저장된 데이터는 사용자로부터 인가된 외부 입력에 의해 삭제, 수정 등의 관리가 이루어질 수 있다.
- [113]
- [114] 도 2는 본 발명의 일 실시예에 따른 검출 모델 또는 구분 모델을 구성하는 신경망 모델의 구조를 나타낸 개략도이다.
- [115] 도 2를 참조하면, 본 발명의 일 실시예에 따른 검체 내 엑소좀 단백질의 변이를 분석하기 위한 검출 모델 또는 구분 모델은 신경망 모델 (300)일 수 있다.
- [116] 신경망 모델 (300)은, 출력의 오류를 최소화하는 방향으로 학습될 수 있다. 예를 들어, 지도 학습의 경우 신경망의 출력과 학습 데이터의 가이드 라벨을 비교하여 오류를 계산하고, 비지도 학습의 경우 신경망의 출력과 학습 데이터를 비교하여 오류를 계산하며, 계산된 오류는 신경망에서 출력 계층으로부터 입력 계층으로 역전파 (backpropagation) 되며, 역전파에 따라 신경망의 노드를 연결하는 각 에지의 가중치 또는 바이어스를 변화시키며, 출력의 오류가 최소화되는 가중치 및 바이어스가 결정될 때까지 학습이 진행될 수 있다. 각 학습에 따른 가중치 또는 바이어스의 변화의 정도는 학습률 (learning rate)에 따라 결정될 수 있다.
- [117] 신경망 모델 (300)은, SERS 신호 데이터를 포함하는 학습 데이터 및 가이드 라벨을 바탕으로 지도 학습되는 것일 수 있다.

- [118] 신경망 모델 (300)이 검출 모델인 경우, 학습 데이터는 야생형 엑소좀 단백질로부터 생성된 SERS 신호 데이터 및 변이 엑소좀 단백질로부터 생성된 SERS 신호 데이터를 포함할 수 있고, 가이드 라벨은 학습 데이터와 대응되는 엑소좀 단백질 변이 여부에 관하여 0과 1, 또는 그 외의 임의의 숫자로 정답을 라벨링한 검출 라벨인 것일 수 있다.
- [119] 신경망 모델 (300)이 구분 모델인 경우, 학습 데이터는 엑소좀 단백질의 변이 차수에 따른 SERS 신호 데이터를 포함하는 차수 학습 데이터인 것일 수 있고, 가이드 라벨은 차수 학습 데이터와 대응되는 변이 차수에 관하여 0과 1, 또는 그 외의 임의의 숫자로 정답을 라벨링한 구분 라벨인 것일 수 있다.
- [120] 학습 데이터는 야생형 엑소좀 단백질 및 변이 엑소좀 단백질을 포함하는 복수의 검체로부터 측정된 복수의 SERS 신호 데이터를 포함할 수 있으며, 학습 데이터에는 신경망 모델 (300)의 학습이 충분히 수행될 수 있는 개수의 SERS 신호 데이터가 포함될 수 있다. 검출 모델에 사용되는 학습 데이터 및 구분 모델에 사용되는 학습 데이터는 동일한 검체로부터 측정된 동일한 SERS 신호 데이터를 포함하지만 각 데이터에 대한 라벨링이 서로 다르게 이루어진 것일 수 있다.
- [121] 신경망 모델 (300)의 학습 과정에서, 학습 데이터에 대한 과적합 (overfitting)이 발생하는 경우 피검 대상으로부터 수득한 검체에 대한 단백질 변이 분석 결과의 정확도가 감소할 수 있으며, 이를 방지하기 위해 입력데이터 및 계층별 입력값에 대한 가중치조정 (regularization), 드롭아웃 (dropout) 또는 배치 정규화 등의 기법이 적용될 수 있다.
- [122] 신경망 모델 (300)에 입력되는 입력데이터는 엑소좀 단백질의 SERS 신호 (200)에 제1변환 (T_1)이 수행되어 생성되는 것일 수 있다.
- [123] SERS 신호 (200)는 피검 대상에서 수득한 검체 내의 엑소좀 단백질로부터 측정되는 것일 수 있다. 엑소좀 단백질의 SERS 신호 측정은 하나의 검체에 대해 1회 또는 복수 회 수행될 수 있고, 이에 따라 하나의 검체로부터 1개 또는 복수 개의 SERS 신호 (200)가 획득될 수 있으며, 예를 들어, 하나의 검체로부터 80개 내지 100개의 SERS 신호 (200)가 획득될 수 있으나, 이에 한정되는 것은 아니다.
- [124] SERS 신호 (200)는 제1변환 (T_1)를 통해 입력데이터로 변환될 수 있다. 제1변환 (T_1)은 SERS 신호 (200)를 신경망 모델 (300)에 입력되기에 적합한 형태로 변환할 수 있는 임의의 방법일 수 있고, 예를 들어, SERS 신호 (200)는 제1변환 (T_1) 과정에서 벡터 데이터의 형태로 수치화된 후 최소-최대 정규화 기법으로 정규화 되어 입력데이터로 변환될 수 있다.
- [125] 입력데이터는 2개 이상의 성분을 가지는 1차원 벡터의 형태일 수 있으며, 예를 들어, 800개의 성분을 가지는 벡터일 수 있으나, 이에 한정되는 것은 아니다.
- [126] 신경망 모델 (300)은, 합성곱 신경망 (310) 및 완전연결 신경망 (320)을 포함할 수 있다.

- [127] 합성곱 신경망 (310)은 하나 이상의 합성곱 단위블록 (311)을 포함할 수 있고, 예를 들어, 5개의 합성곱 단위블록 (311)을 포함할 수 있다.
- [128] 합성곱 단위블록 (311)은 하나 이상의 합성곱 계층을 포함할 수 있고, 배치 정규화 (batch normalization), 활성화 함수 (activation function) 및 풀링 계층을 더 포함할 수 있다.
- [129] 합성곱 계층은 필터를 사용하여 입력값으로부터 특징을 추출할 수 있다. 필터는 2차원 행렬 형태일 수 있고, 필터의 각 성분은 본 발명의 일 실시예에 따른 신경망 모델 (300)의 학습을 통해 최적화되는 매개변수에 해당할 수 있다. 예를 들어, 필터가 3 X 1 크기의 행렬인 경우, 신경망 모델 (300)의 학습 과정에서 해당 필터에 해당하는 데이터의 특징을 추출하여 최적화가 진행될 수 있다.
- [130] 합성곱 계층은 하나 이상의 필터를 포함할 수 있고, 하나 이상의 필터를 통해 입력값으로부터 특징을 추출하는 경우 출력값은 특징 추출에 사용된 필터의 개수와 같은 깊이 (depth)를 가질 수 있다. 예를 들어, 800 x 1 행렬의 입력값에 대해 4개의 필터를 사용하여 연산을 거치는 경우, 출력값은 4의 깊이를 가지게 된다. 따라서, 본 발명의 일 실시예에 따른 합성곱 신경망 (310)에 벡터 형태의 입력데이터가 입력되는 경우, 합성곱 단위블록 (311)에 따른 일련의 연산을 거친 출력값은 3차원 텐서 데이터의 형태를 가진다.
- [131] 합성곱 신경망 (310)의 연산을 모두 거쳐 출력되는 출력값은 제2변환 (T_2)을 거쳐 완전연결 신경망 (320)의 최초 입력값이 될 수 있다.
- [132] 제2변환 (T_2)은 합성곱 신경망 (310)으로부터 출력되는 출력값이 완전연결 신경망 (320)에 입력되기에 적합한 형태로 변환할 수 있는 임의의 방법일 수 있고, 예를 들어, 합성곱 신경망 (310)으로부터 출력되는 2차원 행렬 형태의 출력값은 행렬의 모든 성분이 1차원으로 배열되도록 하는 평탄화 (flattening)을 거쳐 완전연결 신경망 (320)에 입력되는 최초 입력값으로 변환될 수 있다. 이때, 평탄화를 거친 최초 입력값에 대해 배치 정규화가 수행되어 완전연결 신경망 (320)의 학습 효율을 향상시킬 수 있다.
- [133] 완전연결 신경망 (320)은 하나 이상의 완전연결 계층 (321)을 포함할 수 있고, 예를 들어, 4개의 완전연결 계층 (321)을 포함할 수 있다.
- [134] 완전연결 계층 (321)은 입력 계층, 은닉 계층 및 출력 계층을 총칭하는 것일 수 있다. 완전연결 신경망 (320)에 입력된 최초 입력값은 각 계층을 거치며 데이터의 크기가 점차 축소될 수 있으며, 예를 들어, 출력 계층에서 최종적으로 단일 스칼라 값이 출력될 수 있다.
- [135] 완전연결 계층 (321) 중 출력 계층을 제외한 계층 간의 연산에는 sigmoid 함수, tanh 함수, ReLU 함수, ELU 함수 또는 Maxout 함수 등의 활성화 함수에 따른 연산이 포함될 수 있다.
- [136] 본 발명의 일 실시예에서, 완전연결 계층 (321) 중 출력 계층과 인접한 은닉 계층은 최종 벡터 출력 계층 (321a)으로 지칭될 수 있고, 완전연결 계층 (321) 중 출력 계층은 최종 이진 출력 계층 (321b)으로 지칭될 수 있다.

- [137] 신경망 모델 (300)은 검체 내 엑소솜 단백질의 변이에 대한 검출 또는 구분 단계에서, 최종 벡터 출력 계층 (321a)의 출력값 또는 최종 이진 출력 계층 (321b)의 출력값을 이용할 수 있다.
- [138] 최종 벡터 출력 계층 (321a)의 출력값은 2 이상의 길이를 갖는 벡터 형태일 수 있으며, 예를 들어, 128의 길이를 갖는 벡터 데이터일 수 있다.
- [139] 최종 벡터 출력 계층 (321a)의 출력값은 본 발명의 일 실시예에 따른 검출 모델을 통한 엑소솜 단백질의 변이 여부 판단에 이용될 수 있고, 이 경우 최종 벡터 출력 계층 (321a)의 출력값은 상대적 유사도 결정의 기초가 되는 '출력정보'와 상호 교환 가능한 의미로 사용될 수 있다.
- [140] 최종 이진 출력 계층 (321b)의 출력값은 단일 값을 갖는 스칼라 데이터일 수 있으며, 예를 들어, 0 또는 1의 값을 갖는 스칼라 데이터일 수 있다.
- [141] 최종 이진 출력 계층 (321b)의 출력값은 본 발명의 일 실시예에 따른 구분 모델을 통한 변이 엑소솜 단백질의 변이 차수 구분에 이용될 수 있으며, 예를 들어, 구분 모델은 최종 이진 출력 계층 (321b)의 출력값이 0인 경우 변이 엑소솜 단백질에 2차 이상의 변이가 발생하지 않은 것으로 판단할 수 있고, 최종 이진 출력 계층 (321b)의 출력값이 1인 경우 변이 엑소솜 단백질에 2차 이상의 변이가 발생한 것으로 판단할 수 있다.
- [142]
- [143] 도 3은 본 발명의 일 실시예에 따라 검체 내 엑소솜 단백질의 변이를 분석하는 과정을 설명하기 위한 흐름도이다.
- [144] 도 3을 참조하면, 본 발명의 일 실시예에 따라 검체 내 엑소솜 단백질의 변이를 분석하는 과정은 입력데이터 생성 단계 (S100), 변이 엑소솜 단백질 검출 단계 (S200) 및 변이 엑소솜 단백질 구분 단계 (S300)를 포함할 수 있다.
- [145] 입력데이터 생성 단계 (S100)에서, 본 발명의 일 실시예에 따른 컴퓨팅 장치는 검체 내 엑소솜 단백질의 SERS 신호 데이터를 기초로 입력데이터를 생성할 수 있다. 입력데이터는 복수의 SERS 신호 데이터가 통계적으로 처리된 평균 신호를 기초로 생성될 수 있다.
- [146] 입력데이터 중, 피검 대상으로부터 수득한 검체의 엑소솜 단백질이 아닌, 야생형 또는 변이 엑소솜 단백질에서 측정된 SERS 신호를 기초로 생성된 입력데이터는 본 발명의 일 실시예에 따른 검출 모델 및 구분 모델의 학습을 위한 데이터로 사용될 수 있다.
- [147] 변이 엑소솜 단백질 검출 단계 (S200)에서, 컴퓨팅 장치는 미리 학습된 검출 모델을 이용하여, 입력데이터를 기반으로 엑소솜 단백질의 변이 여부를 판단할 수 있다. 이때, 컴퓨팅 장치는 미리 학습된 검출 모델을 이용하여 엑소솜 단백질의 변이 여부에 대한 출력정보를 산출한 후, 출력정보를 기초로 검체 내 엑소솜 단백질과 야생형 엑소솜 단백질 및 변이 엑소솜 단백질의 상대적 유사도를 결정하여 검체 내 엑소솜 단백질이 야생형 단백질 및 변이 단백질 중 어느 것과 더 유사한지 판단할 수 있다.

- [148] 변이 엑소좀 단백질 구분 단계 (S300)에서, 컴퓨팅 장치는 검출 단계에서 엑소좀 단백질이 변이된 것으로 판단된 경우, 미리 학습된 구분 모델을 이용하여 입력데이터를 기반으로 변이 엑소좀 단백질의 변이 차수를 구분할 수 있다. 이때, 컴퓨팅 장치는 검체 내 엑소좀 단백질의 전체 SERS 신호 데이터 중 몇 개의 데이터가 2차 이상의 변이에 따른 데이터인지 판단하여 그 비율을 계산하고, 이를 바탕으로 변이정보를 생성할 수 있다. 컴퓨팅 장치는 2차 이상 변이의 판단 기준인 컷오프 값과 변이정보를 비교하여, 해당 검체 내 엑소좀 단백질에 2차 이상의 변이가 발생한 것인지 구분할 수 있다.
- [149] 변이 엑소좀 단백질 구분 단계 (S300)는 검출 단계 (S200)에 사용된 입력데이터와 동일한 입력데이터를 기초로 수행될 수 있고, 또는 피검 대상에 대한 검출 단계 (S200)가 종료된 후로부터 소정의 기간이 지난 뒤, 동일한 피검 대상으로부터 새로 획득된 검체의 엑소좀 단백질로부터 측정된 SERS 신호로부터 생성된 새로운 입력데이터를 기초로 수행될 수 있다. 따라서, 변이 엑소좀 단백질 구분 단계 (S300)를 통해 현재의 피검 대상의 단백질 변이 차수를 구분할 수 있을 뿐만 아니라, 시간 경과에 따른 피검 대상의 단백질 변이 상태를 모니터링할 수 있다.
- [150] S100 내지 S300의 단계에 따라 검체 내 엑소좀 단백질의 변이를 분석하는 과정의 모식도를 도 4에 나타내었다.
- [151] 본 발명의 일 실시예에 따른 컴퓨팅 장치 (100)가 S100 내지 S300의 단계를 포함하는 동작을 수행하여 검체 내 엑소좀 단백질의 변이를 분석할 수 있도록 하는, 컴퓨터 프로그램은 저장 매체에 저장되어 제공될 수 있다.
- [152] 컴퓨터 프로그램은 일련의 코드 또는 알고리즘으로 구성될 수 있으며, 컴퓨터 프로그램의 구성에 따라 장치가 동작하도록, 독립적으로 또는 결합적으로 프로그램 처리 장치, 예를 들어, 컴퓨팅 장치 (100)를 명령하거나, 처리 장치에 의해 해석될 수 있다. 컴퓨터 프로그램은 처리 장치를 명령하거나 처리 장치에 의해 해석되기 위하여, 임의의 유형의 기계, 구성요소, 물리적 장치, 컴퓨터, 또는 저장 매체를 통해 구체화될 수 있다. 또한, 컴퓨터 프로그램은 물리적 형태가 없는 매체, 예를 들어 네트워크로 연결된 컴퓨터 시스템, 또는 클라우드 (cloud) 서버 등을 통해 분산된 방법으로 저장되거나 실행될 수 있다.
- [153] 저장 매체는 컴퓨터 프로그램을 계속 저장하거나, 실행 또는 다운로드를 위해 임시 저장하는 것일 수 있다. 저장 매체는 단일 또는 수 개의 하드웨어가 결합된 형태의 다양한 기록수단 또는 저장수단일 수 있는데, 어떤 컴퓨터 시스템에 직접 접속되는 매체에 한정되지 않고, 네트워크 상에 분산되어 존재하는 것일 수도 있다. 예를 들어, 저장 매체는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD와 같은 광기록 매체, 플롭티컬 디스크 (floptical disk)와 같은 자기-광 매체 (magneto-optical medium), 및 ROM, RAM, 플래시 메모리 등을 포함하여 컴퓨터 프로그램이 저장되도록 구성된 것일 수 있다.
- [154]

- [155] 이하, 본 발명을 하기의 실시예에 의하여 더욱 상세히 설명한다. 그러나 이들 실시예는 본 발명을 예시하기 위한 것일 뿐이며, 본 발명의 범위가 이들 실시예에 의하여 한정되는 것은 아니다.
- [156]
- [157] 실시예 1. 야생형 및 변이 세포 유래 엑소솜 단백질의 SERS 신호 획득
- [158] 엑소솜 단백질의 SERS 신호를 검출하기 위해 표면에 100 nm 구형 금나노입자가 균일하게 배열되는 SAM (self-assembled monolayer; SAM) 구조를 이용하였다.
- [159] 구체적으로, 세척된 커버글라스 표면에 SAM 구조를 유도할 수 있도록 APTES 처리를 한 뒤 목표 단백질에 특이적으로 결합하는 항체를 부착하고 BSA를 처리하여 비특이적 결합을 방지하였다. 그 후 엑소솜을 용해한 샘플을 커버글라스 표면에 반응시켜 목표 단백질이 항체에 결합하도록 유도하였다. 목표 단백질이 부착된 커버글라스에 농축된 금 나노입자를 뿌린 후 원심분리하여 커버글라스에 균일하게 나노입자가 내려앉도록 하고 건조를 진행하였다. 건조가 완전히 이루어진 후, 목표 단백질이 부착된 커버글라스에 대해 SERS 신호 측정을 진행하였고, 그 결과를 도 5에 나타내었다.
- [160] 이때, 야생형 세포 (A549), 1차 변이 세포 (H838: L858R, HCC82: 19del) 및 2차 변이 세포 (H1975: L858R+T790M, HCC827: 19del+T790M) 유래 엑소솜 샘플에서 각 샘플당 300개의 SERS 신호 데이터 세트를 획득하였고, 각 샘플에서 획득한 신호 데이터 세트를 통계적으로 분석하여 평균 데이터 및 표준편차 데이터를 얻음으로써 샘플별 평균 신호를 형성하였다. 신호 데이터 세트 획득 과정의 모식도를 도 6에 나타내었다.
- [161]
- [162] 실시예 2. 획득한 엑소솜 단백질 SERS 신호를 통한 인공지능 학습
- [163] CNN 모델은 공간적으로 인접한 신호에 대한 상관관계를 추출하기 때문에 전체적인 신호의 추상화된 특징을 도출해낸다. CNN 모델은 기본적으로 2차원의 이미지 분석을 위해 개발되었지만, 본 발명은 1차원 데이터인 라만 신호 분석에 이를 응용하기 때문에, CNN 모델을 다소 간소화하고 수정하였다.
- [164] 그리고, 세포 유래 야생형 엑소솜 단백질 신호와 변이 엑소솜 단백질의 신호를 높은 정확도와 낮은 손실값으로 분류할 수 있고, 세포뿐만 아니라 혈장 단백질의 신호도 잘 분류할 수 있는 모델을 구축하기 위해 다양한 파라미터 및 구조를 최적화하는 과정을 거쳤다.
- [165] 구체적으로, 본 모델의 학습에 사용된 엑소솜 단백질 SERS 신호는 최소최대 정규화 (min-max normalization)을 통해 전처리되었고, $517 \sim 1,800 \text{ cm}^{-1}$ 범위의 라만 스펙트럼을 데이터로 사용하였다.
- [166] 또한, 본 모델의 구조 말단에 있는 완전연결 레이어 (fc 128)는 128개의 값에서 1개의 결괏값으로 연결되는 구조를 가지고 있으며, 이러한 구조를 기반으로 최종적으로 야생형 엑소솜 단백질과 변이 엑소솜 단백질의 SERS 신호 데이터를 분류하였다.

- [167] 이때, 단백질 변이 진단을 위해 체액 내의 엑소솜 등의 생체유래물질을 이용하는 경우의 가장 큰 한계점인 야생형 단백질의 혼재로 인한 낮은 농도의 타겟 단백질 문제를 극복하고자, 각 신호마다 부여되는 최종 예측 도출 값으로서 하나의 신호가 아닌 여러 개의 신호의 평균 도출 값을 이용하였으며, 이를 통해 샘플 내에 타겟 단백질이 낮은 농도로 존재함에도 불구하고 샘플의 전체적인 특징을 추출해내어 해당 샘플 내의 변이 유무를 판단하고자 하였다.
- [168] 본 모델의 훈련에 있어 검출 모델 및 구분 모델 각각의 훈련 및 검증시의 학습 정확도, 및 학습 반복에 따른 손실값의 추이를 측정하여 도 7 및 도 8에 나타내었다.
- [169] 학습 결과, 검출 모델과 구분 모델 모두 단일 SERS 신호를 학습에 사용했을 때 68 ~ 70%의 정확도를 보였지만, 한 샘플 당 80개 이상의 신호 도출값을 통계적으로 분석한 경우 각각 99%, 84% 이상의 정확도를 보여, 모델 학습에 신호의 평균 값을 사용한 경우 더 높은 정확도를 확보할 수 있음을 확인하였다.
- [170]
- [171] 실시예 3. 피검 대상 혈액 유래 엑소솜 단백질의 분석
- [172] 3-1. 엑소솜 단백질의 변이 여부 판단
- [173] 변이 환자 혈액 유래 엑소솜 단백질은 야생형 단백질과 혼재되어있는 상태이기 때문에 세포 유래 엑소솜 단백질을 분류하듯이 이진 분류되기 어렵다. 따라서 본 발명에서는 최종 이진 출력 레이어가 아니라 바로 전 레이어인 최종 벡터 출력 레이어의 정보를 이용하였으며, 이때 최종 벡터 출력 레이어는 앞단의 레이어들을 모두 지난 뒤 도출되는 값으로 128개의 숫자값을 가지는 레이어이다. 변이 엑소솜 단백질을 검출하는 과정의 모식도를 도 9에 나타내었다.
- [174] 구체적으로, 실시예 1과 같은 방법으로 피검 대상의 혈액 내 엑소솜 단백질의 SERS 신호를 측정하고, 이를 기반으로 생성된 입력데이터를 학습된 딥러닝 모델에 입력하여 최종 벡터 출력 레이어의 128개 값을 도출하였다. 그 후, 피검 대상으로부터 도출된 최종 벡터 출력 레이어의 값을 야생형 세포 및 변이 세포의 엑소솜 단백질에서 나타난 최종 벡터 출력 레이어의 값들과 비교하여 어떤 신호와 더 유사한지 분석하였다.
- [175] 이때, 상대적 유사도 척도를 바탕으로 피검 대상의 최종 벡터 출력 레이어와 야생형 세포 및 변이 세포의 엑소솜 단백질의 최종 벡터 출력 레이어 간의 유사도를 분석하였고, 상대적 유사도 척도는 '야생형 단백질과의 마할라노비스 거리/변이 단백질과의 마할라노비스 거리'로 나타내었으며, 여기서 마할라노비스 거리는 각 데이터 간의 비유사성을 나타낸다.
- [176] 상대적 유사도를 바탕으로 정상인과 폐암 변이 환자 (L858R, 19del, T790M)를 분류한 결과를 도 10에 나타내었고, 정상인과 폐암 변이 환자의 평균 상대적 유사도 값을 표 1에 나타내었다.
- [177] [표1]

	정상인 (n=23)	L858R (n=10)	19del (n=11)	T790M (n=3)
평균 상대적 유사도(a. u.)	0.263	0.45	0.61	1.66

[178] 분류 결과, 육안으로도 정상인과 폐암 변이 환자가 통계적으로 유의미하게 분류되는 것을 확인하였으며, 다양한 변이 종류에서 모두 유의미하게 분류되는 것을 확인하였다.

[179] 아울러, 본 발명에 따라 학습된 검출 모델의 진단 효과를 검증하기 위해, ROC 곡선을 도 11에 나타내었고, AUC 값을 기존의 문헌 데이터와 비교하여 표 2에 나타내었다.

[180] [표2]

저자	진단 방법	AUC
Paolo Cotzia, et. al.	Plasma sequencing based ctDNA detection	0.63
Lei Wang, et. al.	PCR and SERS based ctDNA detection	0.85
Cheng-Yu Chen, et. al.	CT image-Radiomics based classification	0.89
H. Alvarez, et. al.	dd-PCR based exoDNA mutation detection	0.66-0.85
본 발명	Deep-learning and SERS based classification	0.85

[181] 도 11 및 표 2를 참고할 때, 본 발명의 검출 모델의 AUC 값은 0.85로, CT 이미지-Radiomics을 이용한 진단방법 (Cheng-Yu Chen, et al.) 다음으로 유용성이 있으며, 액체생검 기반의 폐암 진단 방법에서는 최상위의 유용성을 가지는 것으로 나타났다. 결과적으로, 본 발명은 매우 정확한 수준의 진단 방법으로서 실제 진단에 응용될 수 있는 가능성을 내포하는 것으로 판단되었다.

[182]

[183] **3-2. 변이 엑소좀 단백질의 변이 차수 구분**

[184] 1차 변이 환자 혈액 유래 엑소좀 단백질은 야생형 단백질, 1차 변이 단백질이 혼재되어있지만, 2차 변이 환자 유래 엑소좀 단백질은 다량의 야생형 단백질, 1차 변이 단백질 및 소량의 2차 변이 단백질이 아주 소량 혼재되어 있다. 때문에, 실시예 3-1의 방법처럼 전체적인 유사도 값을 통해서 변이를 구분하더라도 2차 변이 단백질의 특징이 잘 담기지 않아 분류가 어렵다.

[185] 따라서, 한 샘플 당 100개의 신호 중 딥러닝 모델을 통해 2차 변이라고 판단된 신호의 개수를 산출하고, 2차 변이라고 판단된 신호의 개수를 전체 신호 개수 (n=100)로 나눔으로써 피검 대상 유래 엑소좀 단백질의 2차 변이 값, 즉 변이정보

를 도출하였다. 변이 엑소좀 단백질을 구분하는 과정의 모식도를 도 12에 나타내었고, 변이 값을 바탕으로 폐암 환자 중 1차 변이 및 2차 변이 환자를 구분한 결과를 도 13에 나타내었다

- [186] 구분 결과, 육안으로도 변이 구분의 컷오프 값을 기준으로 1차 변이 및 2차 변이 환자가 잘 구분되는 것을 확인하였다.
- [187] 아울러, 1차 변이 진단 시점 (t1) 및 2차 변이 진단 시점 (t2)에 있어, 변이정보를 바탕으로 폐암 환자 중 1차 변이 및 2차 변이 환자를 구분한 예측값을 시각화한 결과를 도 14에 나타내고, 1차 변이 진단 시점 (t1) 대비 2차 변이 진단 시점 (t2)에서의 2차 변이 신호 증감 비율을 도 15에 나타내었다. 이때, 1차 변이로 판단되는 신호는 노란색 박스, 2차 변이로 판단되는 신호는 보라색 박스로 시각화하였다.
- [188] 이와 같이 환자의 치료 모니터링을 수행한 결과, 진단 시점에 따라 2차 변이 단백질 개수가 증가하는 것으로 나타나, 본 발명은 새로운 변이 스크리닝 방법으로 응용될 수 있다는 점이 시사되었다.
- [189] [부호의 설명]
- [190] 100: 컴퓨팅 장치
- [191] 200: 엑소좀 단백질의 SERS 신호
- [192] 300: 신경망 모델
- [193] 310: 합성곱 신경망
- [194] 320: 완전연결 신경망
- [195] T₁: 제1변환
- [196] T₂: 제2변환
- [197] S100: 입력데이터 생성 단계
- [198] S200: 변이 엑소좀 단백질 검출 단계
- [199] S300: 변이 엑소좀 단백질 구분 단계

산업상 이용가능성

- [200] 본 발명은 혈액 내 풍부하게 존재하며 바이오마커를 안정적으로 운반하는 엑소좀에 함유된 단백질의 표면증강 라만 산란 (SERS) 신호를 합성곱 신경망 (CNN) 기반의 인공지능 알고리즘을 통해 분류하고, 도출된 분류 결과를 수학적으로 해석하여 엑소좀 단백질의 변이를 검출 및 모니터링할 수 있는 방법 및 장치에 관한 것으로, 본 발명을 통해 기존의 기술들이 가지고 있는 신호 이질성의 한계를 극복할 수 있으며, 이에 따라 혈장 내 엑소좀 단백질에 대한 간편하고 신속한 비침습적인 액체 생검이 가능하다.

청구범위

- [청구항 1] 컴퓨팅 장치에 의해 수행되는, 검체 내 엑소좀 단백질의 변이를 분석하는 방법에 있어서,
 검체 내 엑소좀 단백질의 표면증강 라만 산란 (SERS) 신호 데이터를 기초로 입력데이터를 생성하는 생성 단계; 및
 미리 학습된 검출 모델을 이용하여, 상기 입력데이터를 기반으로 엑소좀 단백질의 변이 여부를 판단하는 검출 단계;
 를 포함하는, 검체 내 엑소좀 단백질의 변이를 분석하는 방법.
- [청구항 2] 제1항에 있어서, 상기 검체 내 엑소좀 단백질의 변이를 분석하는 방법은, 검체 내 엑소좀 단백질로부터 SERS 신호를 측정하는 신호 측정 단계;
 를 포함하는 것인, 검체 내 엑소좀 단백질의 변이를 분석하는 방법.
- [청구항 3] 제1항에 있어서, 상기 생성 단계는,
 복수의 SERS 신호 데이터를 통계적 처리를 통해 평균 신호로 변환하는 통계적 처리 단계;
 를 포함하는 것인, 검체 내 엑소좀 단백질의 변이를 분석하는 방법.
- [청구항 4] 제3항에 있어서, 상기 통계적 처리 단계는,
 80개 내지 100개의 SERS 신호 데이터를 통계적 처리를 통해 평균 신호로 변환하는 것인, 검체 내 엑소좀 단백질의 변이를 분석하는 방법.
- [청구항 5] 제1항에 있어서, 상기 검출 단계는,
 미리 학습된 검출 모델을 이용하여, 엑소좀 단백질의 변이 여부에 대한 출력정보를 산출하는 산출 단계; 및
 상기 출력정보를 기초로, 검체 내 엑소좀 단백질과 야생형 엑소좀 단백질 및 변이 엑소좀 단백질의 상대적 유사도를 결정하는 유사도 결정 단계;
 를 포함하는, 검체 내 엑소좀 단백질의 변이를 분석하는 방법.
- [청구항 6] 제5항에 있어서, 상기 출력정보는 길이가 128인 1차원의 텐서 데이터인 것인, 검체 내 엑소좀 단백질의 변이를 분석하는 방법.
- [청구항 7] 제1항에 있어서, 상기 검출 모델은,
 야생형 엑소좀 단백질로부터 생성된 SERS 신호 데이터 및 변이 엑소좀 단백질로부터 생성된 SERS 신호 데이터를 포함하는 학습 데이터, 및 상기 학습 데이터와 대응되는 엑소좀 단백질 변이 여부에 관한 검출 라벨에 기초하여 지도 학습되며,
 상기 지도 학습은 상기 검출 모델을 이용하여 상기 학습 데이터로부터 생성한 학습 정보, 및 상기 검출 라벨에 기초하여 수행되는 것인, 검체 내 엑소좀 단백질의 변이를 분석하는 방법.
- [청구항 8] 제1항에 있어서, 상기 검체 내 엑소좀 단백질의 변이를 분석하는 방법은,

상기 검출 단계에서 엑소좀 단백질이 변이된 것으로 판단된 경우, 미리 학습된 구분 모델을 이용하여, 상기 입력데이터를 기반으로 변이 엑소좀 단백질의 변이 차수를 구분하는 구분 단계;

를 포함하는 것인, 엑소좀 단백질의 변이를 분석하는 방법.

[청구항 9] 제8항에 있어서, 상기 구분 단계는, 상기 검출 단계에서 엑소좀 단백질이 변이된 것으로 판단된 경우, 상기 검출 단계 종결 이후 소정의 기간이 경과한 후, 동일한 피검 대상으로부터 새로 취득된 검체의 엑소좀 단백질에 대해 수행되는 것인, 엑소좀 단백질의 변이를 분석하는 방법.

[청구항 10] 제8항에 있어서, 상기 구분 단계는, 전체 SERS 신호 데이터 중 2차 이상의 변이가 존재하는 것으로 판단된 SERS 신호 데이터의 비율을 계산하여 변이정보를 생성하는 변이정보 생성 단계;

를 포함하는 것인, 엑소좀 단백질의 변이를 분석하는 방법.

[청구항 11] 제10항에 있어서, 상기 전체 SERS 신호 데이터는 검체 내 엑소좀 단백질로부터 측정된 100개의 SERS 신호를 기반으로 생성되는 것인, 검체 내 엑소좀 단백질의 변이를 분석하는 방법.

[청구항 12] 제8항에 있어서, 상기 구분 모델은, 엑소좀 단백질의 변이 차수에 따른 SERS 신호 데이터를 포함하는 차수 학습 데이터, 및 상기 차수 학습 데이터와 대응되는 변이 차수에 관한 구분 라벨에 기초하여 지도 학습되며, 상기 지도 학습은 상기 구분 모델을 이용하여 상기 차수 학습 데이터로부터 생성한 학습 정보, 및 상기 구분 라벨에 기초하여 수행되는 것인, 검체 내 엑소좀 단백질의 변이를 분석하는 방법.

[청구항 13] 저장 매체에 저장된 컴퓨터 프로그램으로서, 컴퓨터 프로그램은 하나 이상의 프로세서에서 실행되는 경우, 검체 내 엑소좀 단백질의 변이를 분석하기 위한 동작들을 수행하며, 상기 동작들은 하기를 포함하는 것인, 저장 매체에 저장된 컴퓨터 프로그램:

검체 내 엑소좀 단백질의 SERS 신호 데이터를 기초로 입력데이터를 생성하는 동작; 및

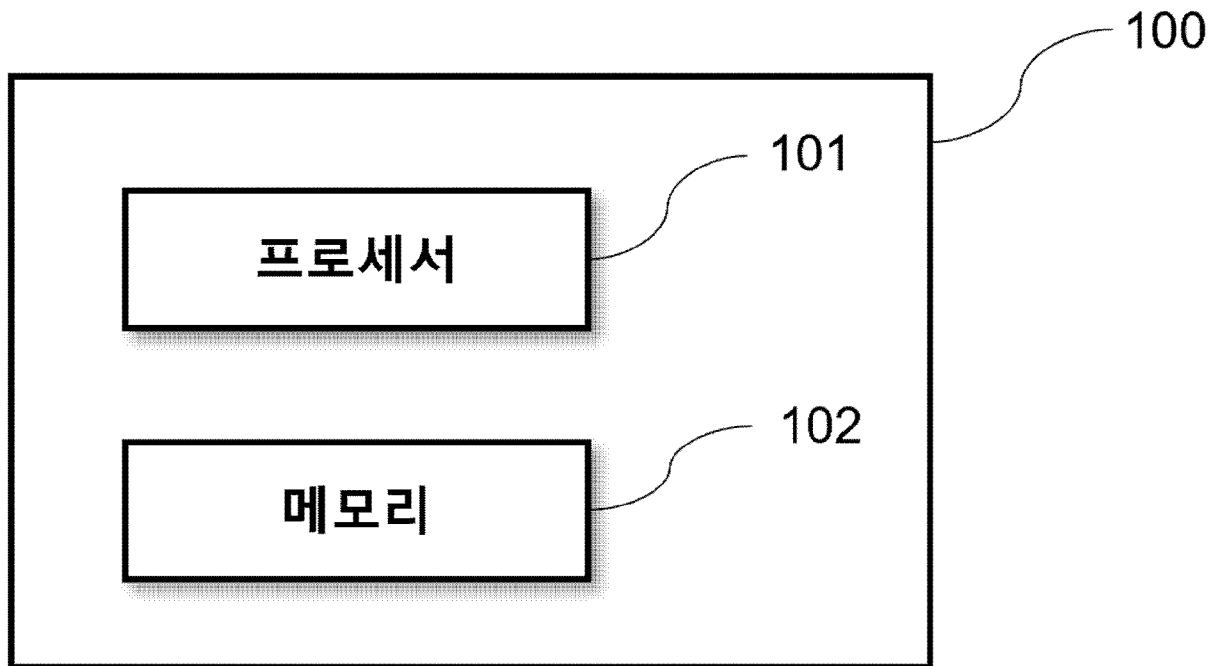
미리 학습된 검출 모델을 이용하여, 상기 입력데이터를 기반으로 엑소좀 단백질의 변이 여부를 판단하는 동작.

[청구항 14] 검체 내 엑소좀 단백질의 변이를 분석하기 위한 컴퓨팅 장치로서, 하나 이상의 코어를 포함하는 프로세서; 및 메모리;를 포함하고,

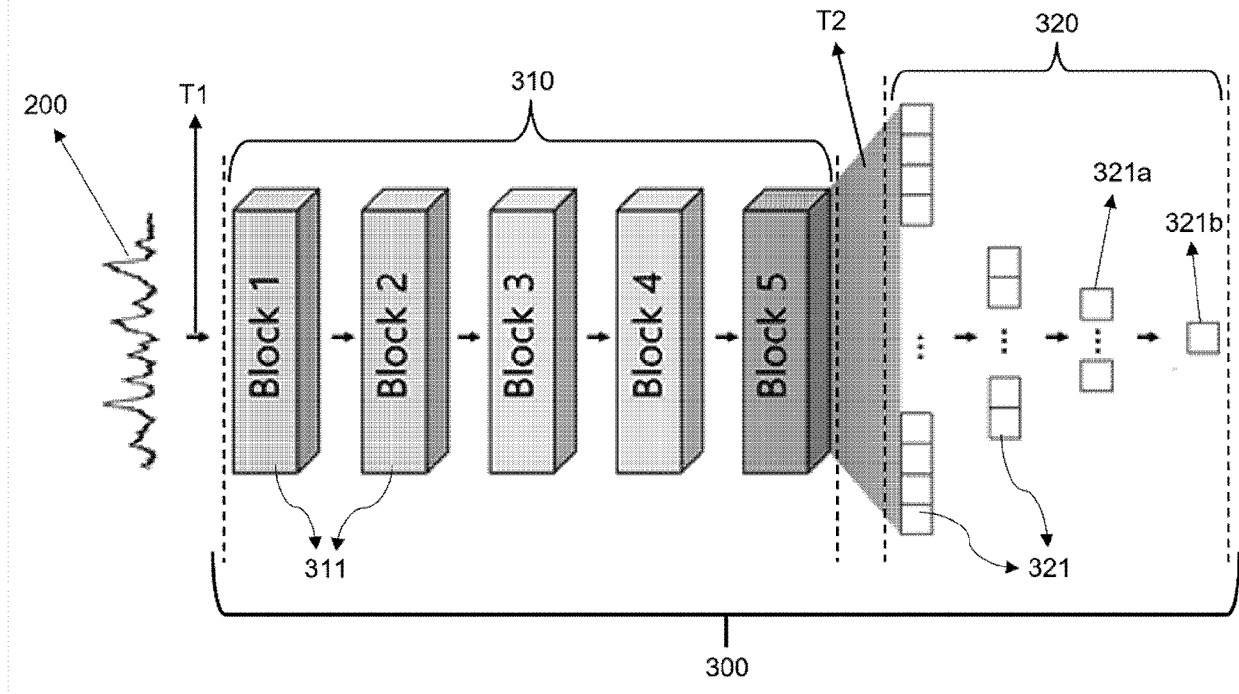
상기 프로세서는, 검체 내 엑소좀 단백질의 SERS 신호 데이터를 기초로 입력데이터를 생성하고, 및

미리 학습된 검출 모델을 이용하여, 상기 입력데이터를 기반으로 엑소좀 단백질의 변이 여부를 판단하는 것인,
컴퓨팅 장치.

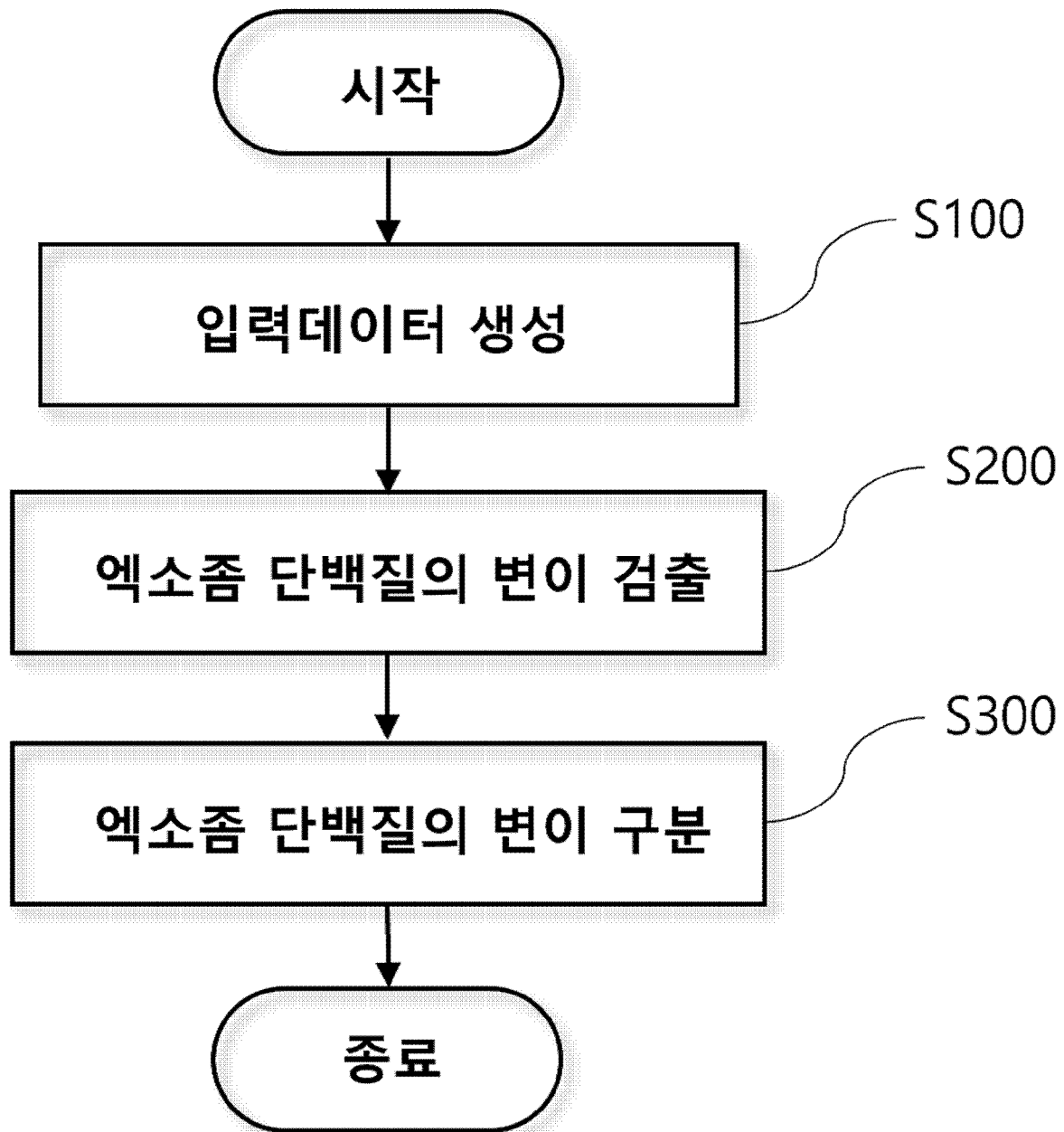
[도1]



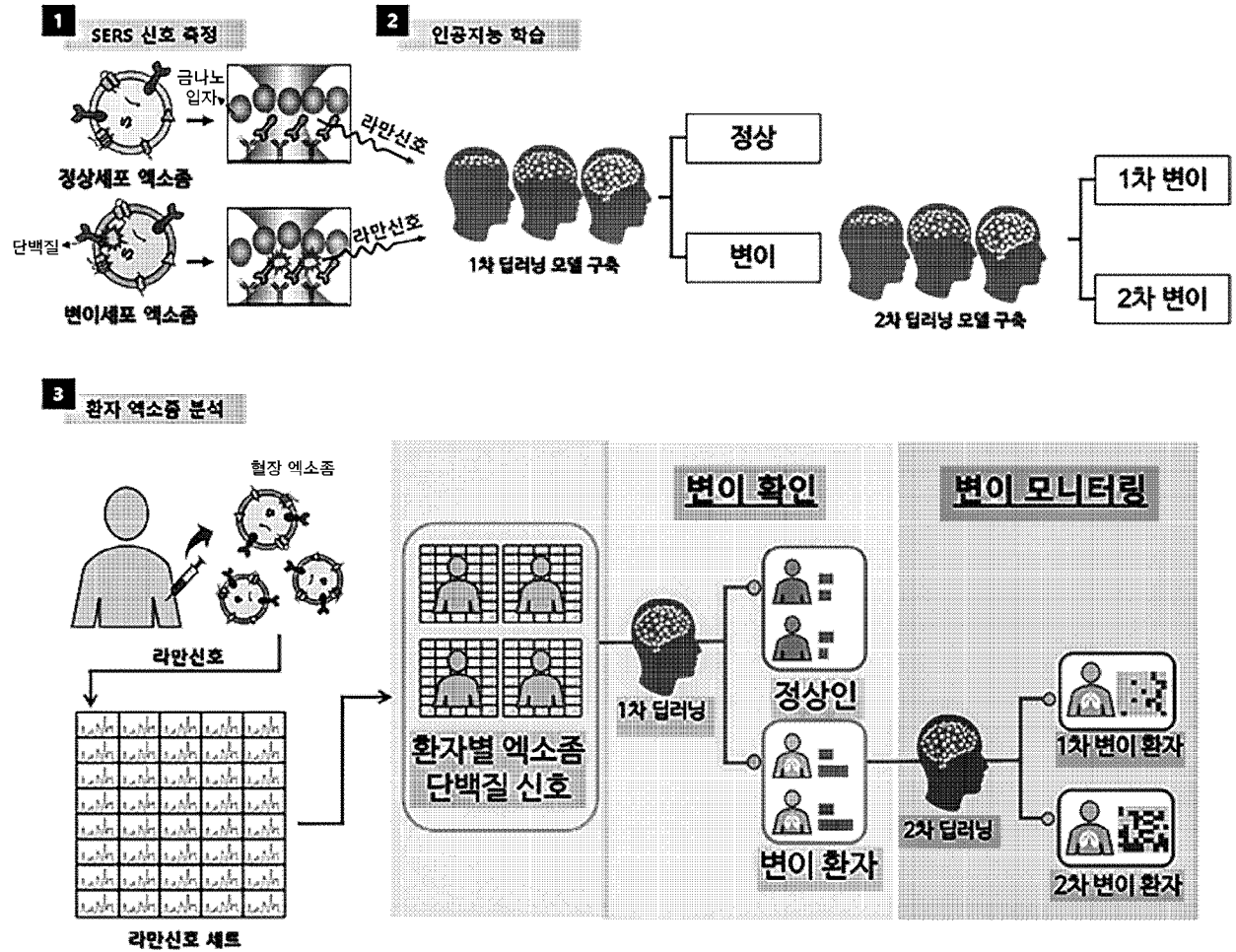
[도2]



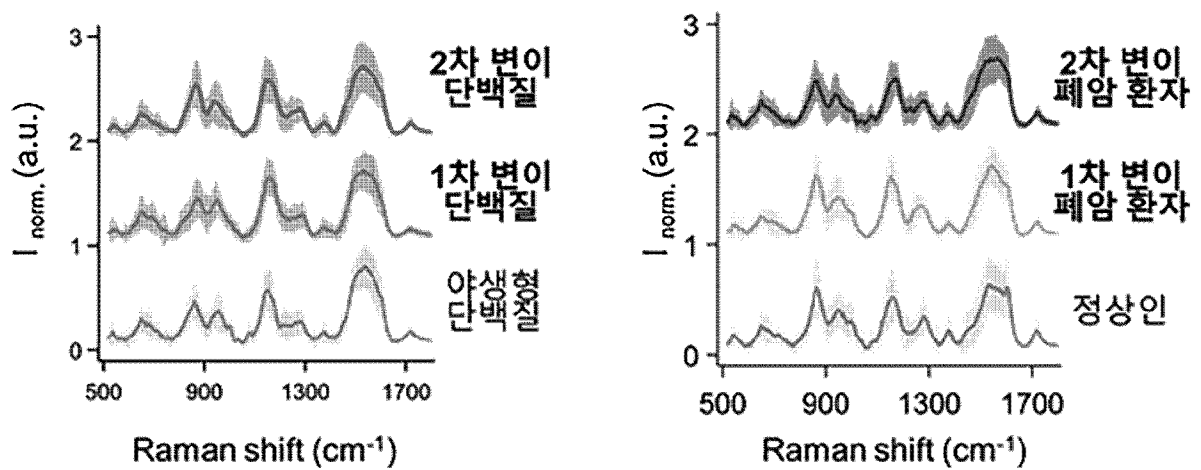
[도3]



[도4]



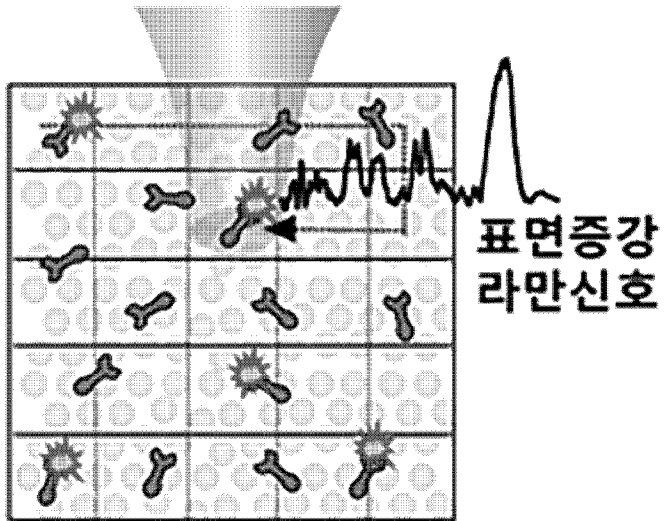
[도5]



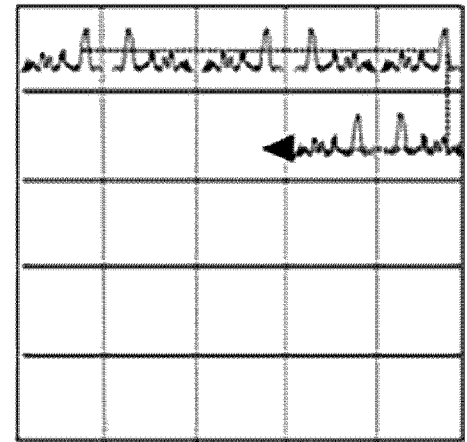
[도6]

1

기판 예상 모식도

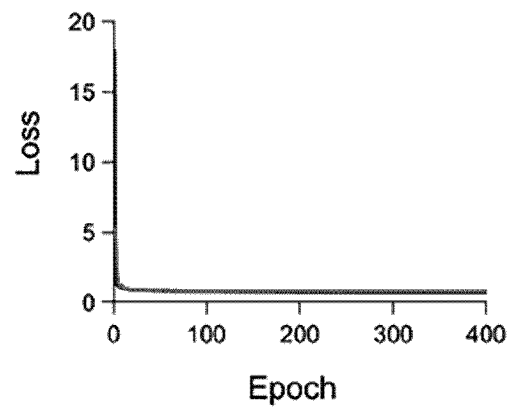
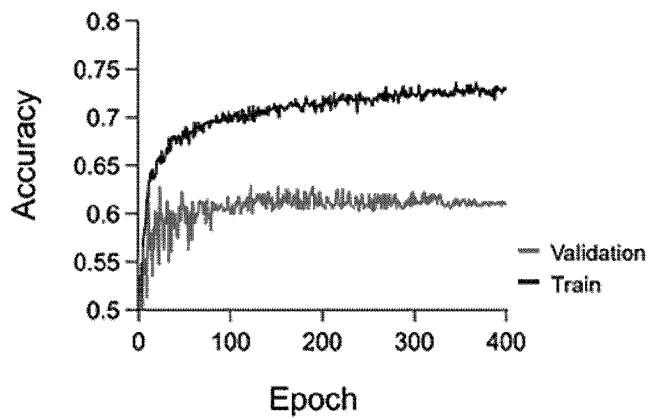


금나노기판

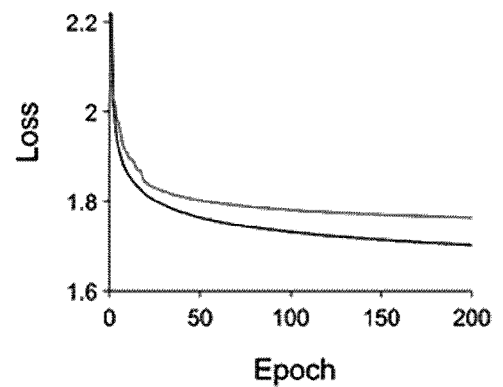
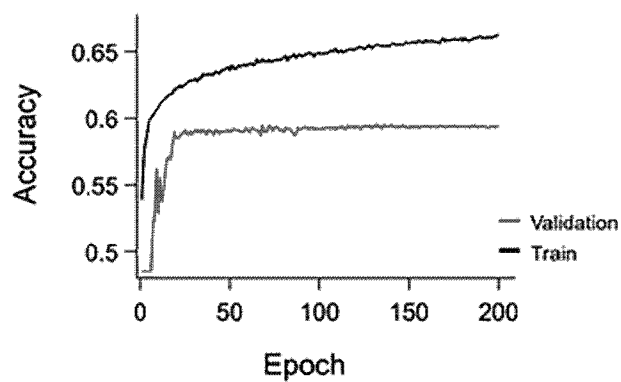


라만 신호 데이터 셋

[도7]

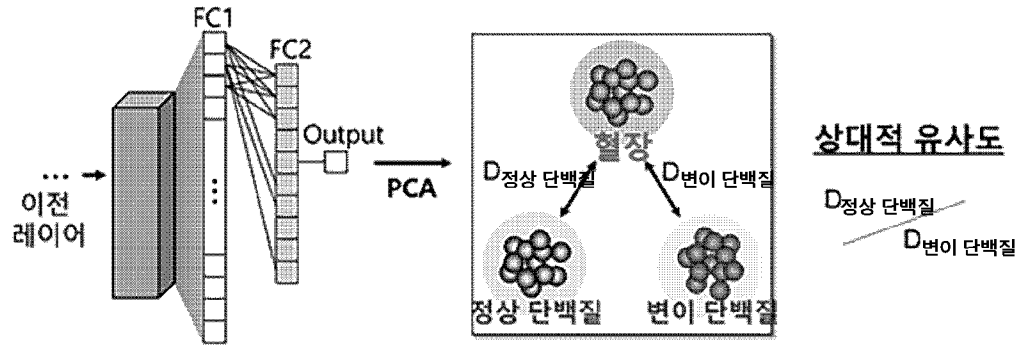


[도8]

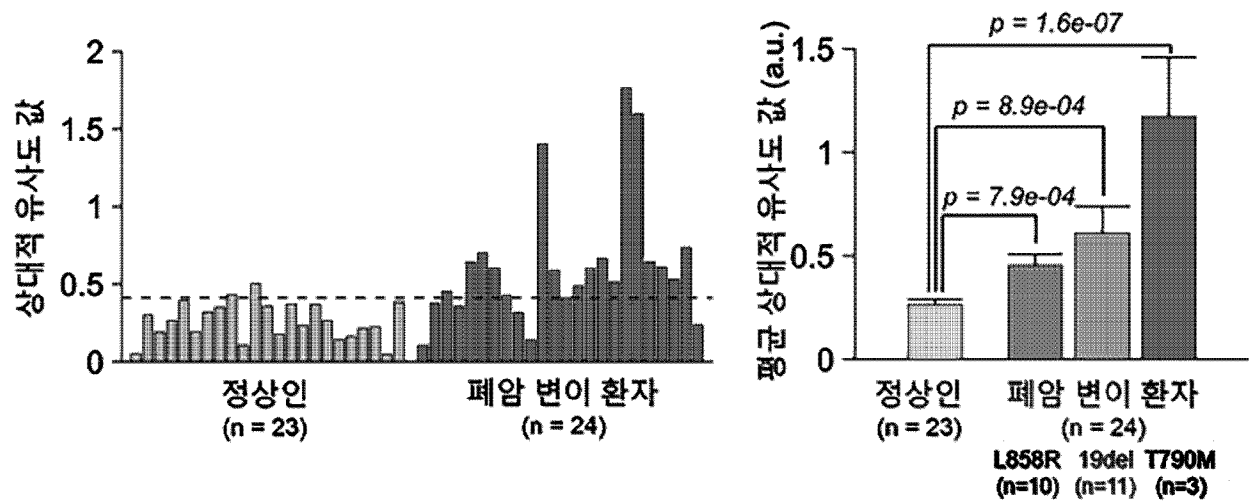


[도9]

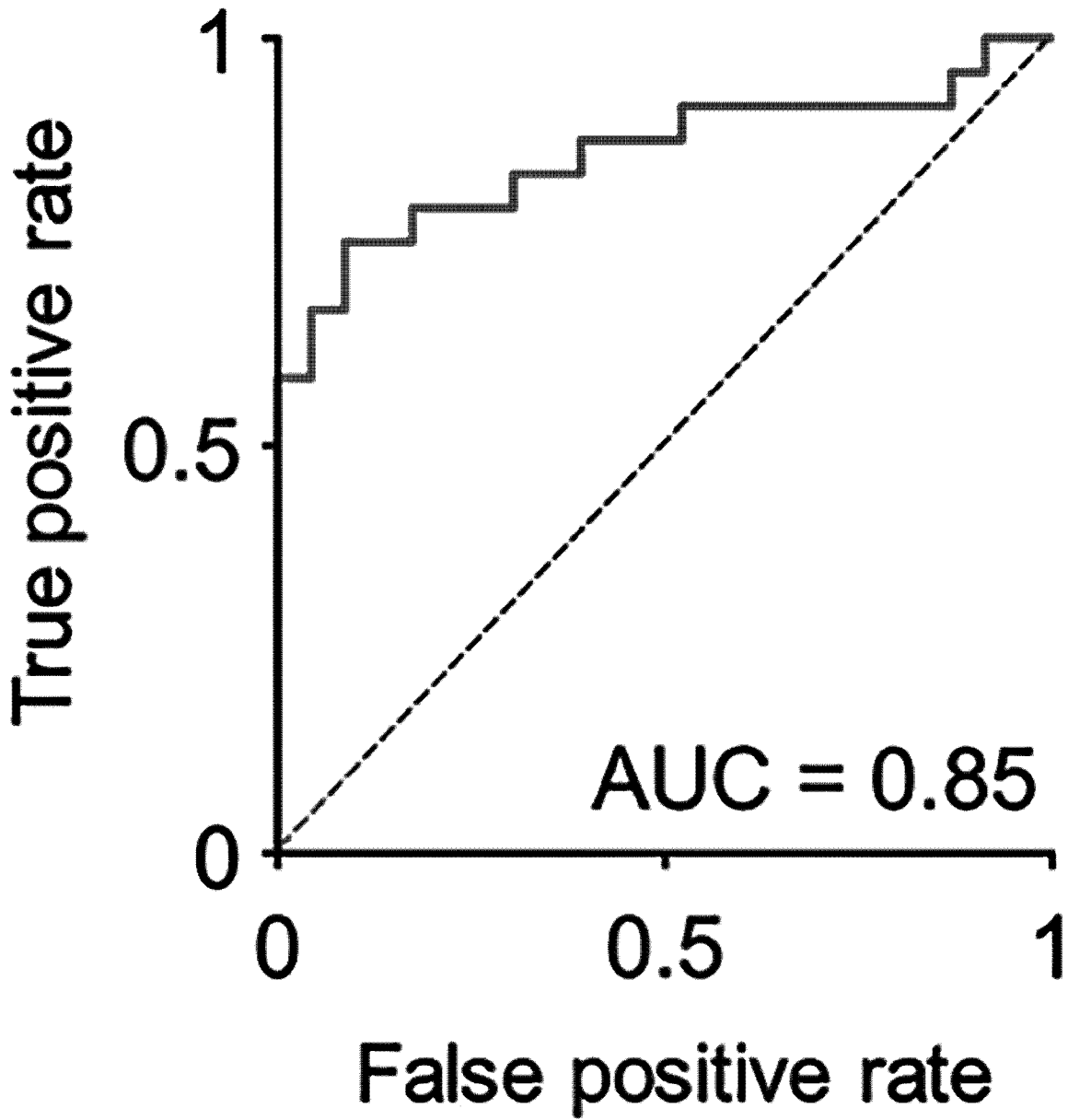
1 변이 세포와의 상대적 유사도 값 도출 모식도



[도10]

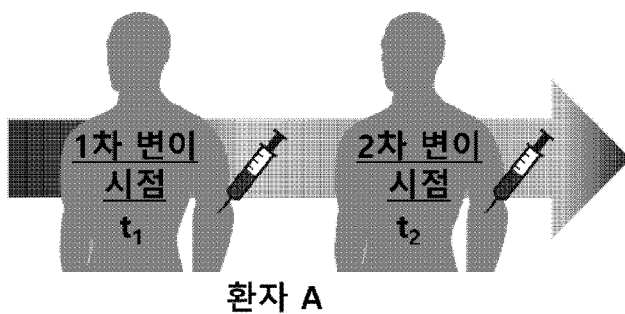


[도11]

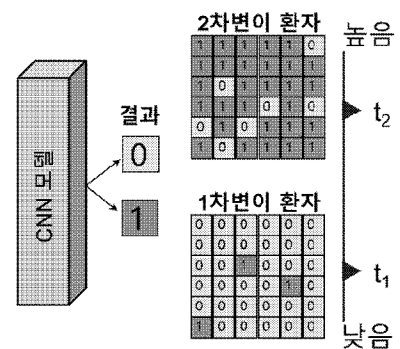


[도12]

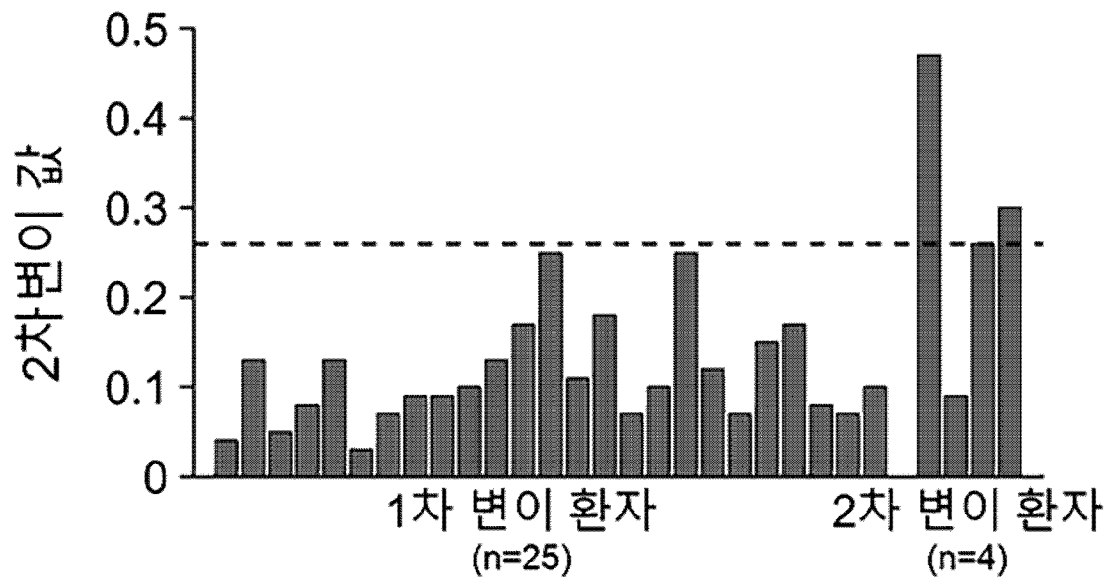
1 장기 치료에 따른 저항 변이 발생



2 환자 별 2차변이 예측값 도출 모식도



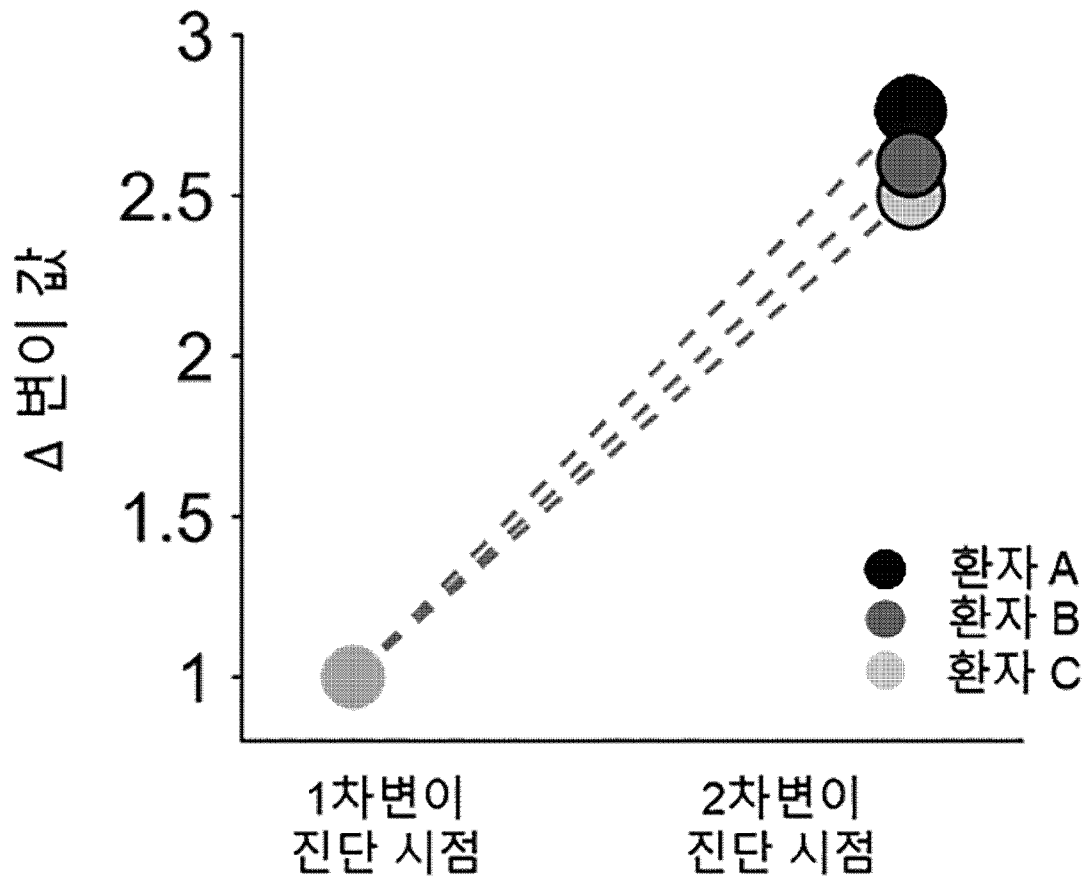
[도13]



[도14]

		환자 A	환자 B	환자 C
2차변이 값 = $\frac{\#}{n}$	t ₁			
		0.17	0.1	0.12
	t ₂			
		0.47	0.26	0.3

[도15]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2024/001171

A. CLASSIFICATION OF SUBJECT MATTER**G16B 20/20**(2019.01)i; **G16B 40/20**(2019.01)i; **G06N 3/0464**(2023.01)i; **G16B 45/00**(2019.01)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)

G16B 20/20(2019.01); A61B 5/00(2006.01); G01J 3/02(2006.01); G01N 21/65(2006.01); G01N 33/483(2006.01)

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) & keywords: 엑소좀(exosomes), 미세소포체(extracellular vesicles), 단백질(protein), 변이(mutation), 표면증강 라만 산란(Surface Enhanced Raman Spectroscopy, SERS), 액체생검(liquid biopsy), 인공지능(artificial intelligence, AI), 기계 학습(machine learning), 지도 학습(supervised learning), 텐서 데이터(tensor data), 마할라노비스 거리(Mahalanobis distance), 합성곱 신경망(convolutional neural network, CNN)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KR 10-2020-0108938 A (EXOPERT CORPORATION) 22 September 2020 (2020-09-22) See claims 1-7; paragraphs [0028]-[0034], [0046]-[0050], [0070]-[0073] and [0103]; and figures 6 and 9.	1-7,13,14
A		8-12
Y	LYU, N. et al. SERS characterization of colorectal cancer cell surface markers upon anti-EGFR treatment. Exploration. 09 May 2022, vol. 2, no. 3, article no. 20210176, inner pp. 1-10. See abstract; inner pages 1-4; and figures 1-3.	1-7,13,14
A	SHIN, H. et al. Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes. ACS Nano. 2020, vol. 14, no. 5, pp. 5435-5444. See abstract; and pages 5435-5442.	1-14
A	KR 10-2018-0010280 A (KOREA UNIVERSITY RESEARCH AND BUSINESS FOUNDATION) 30 January 2018 (2018-01-30) See abstract; and claims 1 and 2.	1-14



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

“D” document cited by the applicant in the international application

“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

02 May 2024

Date of mailing of the international search report

02 May 2024

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2021-0372933 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 02 December 2021 (2021-12-02) See abstract; and claims 1-4.	1-14
PX	KIM, S. et al. Plasma Exosome Analysis for Protein Mutation Identification Using a Combination of Raman Spectroscopy and Deep Learning. ACS Sensors. 06 June 2023, vol. 8, no. 6, pp. 2391-2400. See abstract; and pages 2391-2397.	1-14

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2024/001171

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
KR	10-2020-0108938	A	22 September 2020	CN	113194820	A	30 July 2021
				EP	3936031	A1	12 January 2022
				EP	3936031	A4	07 December 2022
				KR	10-2291105	B1	23 August 2021
				US	2022-0022816	A1	27 January 2022
				WO	2020-180003	A1	10 September 2020
KR	10-2018-0010280	A	30 January 2018	KR	10-2017-0115907	A	18 October 2017
				WO	2017-175928	A1	12 October 2017
US	2021-0372933	A1	02 December 2021	US	11680907	B2	20 June 2023
				US	2022-412892	A1	29 December 2022
				WO	2019-140305	A1	18 July 2019

A. 발명이 속하는 기술분류(국제특허분류(IPC)) G16B 20/20(2019.01)i; G16B 40/20(2019.01)i; G06N 3/0464(2023.01)i; G16B 45/00(2019.01)i		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) G16B 20/20(2019.01); A61B 5/00(2006.01); G01J 3/02(2006.01); G01N 21/65(2006.01); G01N 33/483(2006.01) 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) & 키워드: 엑소좀(exosomes), 미세소포체(extracellular vesicles), 단백질(protein), 변이(mutation), 표면증강 라만 산란(Surface Enhanced Raman Spectroscopy, SERS), 액체생검(liquid biopsy), 인공 지능(artificial intelligence, AI), 기계 학습(machine learning), 지도 학습(supervised learning), 텐서 데이터(tensor data), 마할라노비스 거리(Mahalanobis distance), 합성곱 신경망(convolutional neural network, CNN)		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y A	KR 10-2020-0108938 A (주식회사 엑소퍼트) 2020.09.22 청구항 1-7; 단락 [0028]-[0034], [0046]-[0050], [0070]-[0073], [0103]; 및 도 6, 9	1-7,13,14 8-12
Y	LYU, N. 등, "SERS characterization of colorectal cancer cell surface markers upon anti-EGFR treatment", Exploration, 2022년 05월 09일, 2권, 3호, 기사번호 20210176, 내부페이지 1-10 요약; 내부페이지 1-4; 및 도 1-3	1-7,13,14
A	SHIN, H. 등, "Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes", ACS Nano, 2020년, 14권, 5호, 페이지 5435-5444 요약; 및 페이지 5435-5442	1-14
A	KR 10-2018-0010280 A (고려대학교 산학협력단) 2018.01.30 요약; 및 청구항 1, 2	1-14
☒ 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 "D" 본 국제출원에서 출원인이 인용한 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 2024년05월02일(02.05.2024)		국제조사보고서 발송일 2024년05월02일(02.05.2024)
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578		심사관 허주형 전화번호 +82-42-481-5373

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	US 2021-0372933 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 2021.12.02 요약; 및 청구항 1-4	1-14
PX	KIM, S. 등, "Plasma Exosome Analysis for Protein Mutation Identification Using a Combination of Raman Spectroscopy and Deep Learning", ACS Sensors, 2023년 06월 06일, 8권, 6호, 페이지 2391-2400 요약; 및 페이지 2391-2397	1-14

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2020-0108938 A	2020/09/22	CN 113194820 A	2021/07/30
		EP 3936031 A1	2022/01/12
		EP 3936031 A4	2022/12/07
		KR 10-2291105 B1	2021/08/23
		US 2022-0022816 A1	2022/01/27
		WO 2020-180003 A1	2020/09/10
KR 10-2018-0010280 A	2018/01/30	KR 10-2017-0115907 A	2017/10/18
		WO 2017-175928 A1	2017/10/12
US 2021-0372933 A1	2021/12/02	US 11680907 B2	2023/06/20
		US 2022-412892 A1	2022/12/29
		WO 2019-140305 A1	2019/07/18