

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국



(10) 국제공개번호

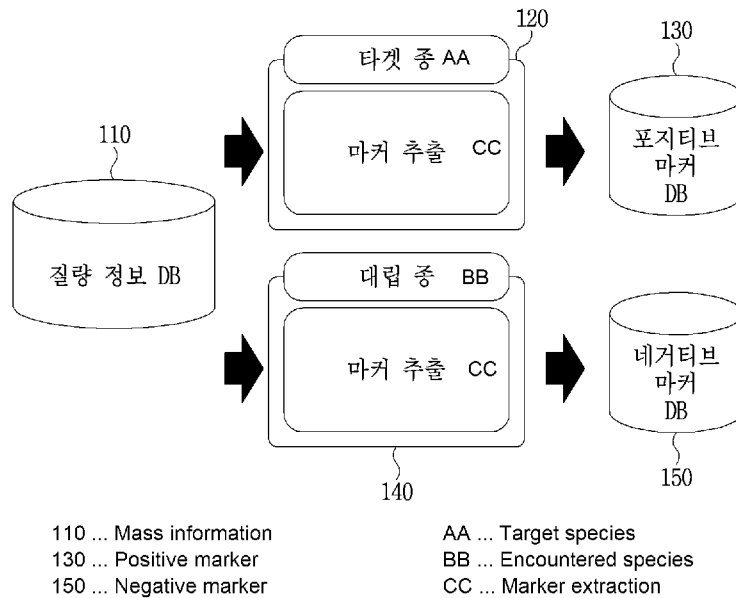
(43) 국제공개일
2018년 12월 27일 (27.12.2018) WIPO | PCT

WO 2018/236120 A1

- (51) 국제특허분류: *G01N 33/68* (2006.01) *G06N 3/08* (2006.01) *G06F 19/22* (2011.01)
- (21) 국제출원번호: PCT/KR2018/006892
- (22) 국제출원일: 2018년 6월 19일 (19.06.2018)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보: 62/524,023 2017년 6월 23일 (23.06.2017) US
- (71) 출원인: 주식회사 에이엠아이티 (AMIT, INC.) [KR/KR]; 13494 경기도 성남시 분당구 대왕판교로 660, A-908, Gyeonggi-do (KR).
- (72) 발명자: 이종서 (LEE, Jongseo); 16509 경기도 수원시 영통구 예류타운로 35, 5103-1202, Gyeonggi-do (KR). 김성국 (KIM, Songkuk); 13542 경기도 성남시 분당구 판교원로186, 403-301, Gyeonggi-do (KR). 조웅준 (JO, EungJoon); 07675 뉴저지주 올드타운 스위트 3 올드타운로드 빌딩 1, 180, New Jersey (US).
- (74) 대리인: 최윤서 등 (CHOE, Yun Seo et al.); 06731 서울시 서초구 서운로 26-1, 501호 윤특허법률사무소, 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,

(54) Title: METHOD AND DEVICE FOR IDENTIFYING QUASISPECIES BY USING NEGATIVE MARKER

(54) 발명의 명칭: 네거티브 마커를 이용한 유사 종 식별 방법 및 장치



(57) Abstract: The present disclosure relates to a method and a device for identifying a quasispecies and, more particularly, to a method and a device for identifying a quasispecies on the basis of machine learning using a negative marker. A method for identifying a quasispecies according to an embodiment of the present disclosure may comprise the steps of: extracting first mass information regarding an inputted sample; classifying the inputted sample on the basis of the first mass information, at least by using a machine learning model based on a negative marker; and identifying the species regarding the inputted sample on the basis of the result of classification.

(57) 요약서: 본 개시는 유사 종 식별 방법 및 장치에 대한 것이며, 구체적으로는 네거티브 마커를 이용한 머신 러닝에 기초하여 유사 종을 식별하는 방법 및 장치에 대한 것이다. 본 개시의 일 실시예에 따른 유사 종을 식별하는 방법은, 입력된 샘플에 대한 제 1 질량 정보를 추출하는 단계; 상기 제 1 질량 정보에 기초하여, 적어도 네거티브 마커에 기반한 머신 러닝 모델을 이용하여 상기 입력된 샘플을 분류하는 단계; 및 상기 분류 결과에 기초하여 상기 입력된 샘플에 대한 종을 식별하는 단계를 포함할 수 있다.

LR, LS, MW, MZ, NA, RW, 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, 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))

명세서

발명의 명칭: 네거티브 마커를 이용한 유사 종 식별 방법 및 장치 기술분야

[1] 관련 출원에 대한 상호-참조(CROSS-REFERENCE TO RELATED APPLICATIONS)

[2] 본 출원이 우선권을 주장하는 2017년 6월 23일자 출원된 미국 임시출원번호 제62/524,023호는 그 전체가 본 개시에 참조로써 통합된다(The present application claims priority to U.S. Provisional Patent Application No. 62/524,023 filed on June 23, 2017, which is hereby incorporated by reference in its entirety).

[3] 본 개시는 유사 종 식별 방법 및 장치에 대한 것이며, 구체적으로는 네거티브 마커를 이용한 머신 러닝에 기초하여 유사 종을 식별하는 방법 및 장치에 대한 것이다.

배경기술

[4] 질량 스펙트럼 분석 방식은 대상의 질량 조성을 식별하기 위해 널리 사용되고 있다. 예를 들어, 미지의 미생물에 대해서 추출된 질량 정보에 기초하여 선택된 마커(marker)를 적용함으로써 해당 미생물을 식별할 수 있다. 마커는 미생물을 고유하게 식별할 수 있는 특성을 의미한다. 또한, 추출된 질량 조성 정보와 머신 러닝 기법을 조합함으로써 미생물 식별 성능을 개선할 수 있다.

[5] 이러한 질량 스펙트럼 분석 방식에 의하더라도, 유사한 미생물 종들의 질량 스펙트럼 패턴은 서로 매우 유사하기 때문에, 종래의 방법을 통해서는 유사한 미생물 종들을 정확하게 식별 또는 구별하는 것이 어렵다. 따라서, 유사 종들간의 식별 성능을 개선하기 위한 방안이 요구된다.

발명의 상세한 설명

기술적 과제

[6] 본 개시의 기술적 과제는 유사 종들간의 식별 성능을 개선하기 위한 방법 및 장치를 제공하는 것이다.

[7] 본 개시의 추가적인 기술적 과제는 머신 러닝 기법에 무관하게 미생물 식별 성능을 개선하기 위한 방법 및 장치를 제공하는 것이다.

[8] 본 개시의 추가적인 기술적 과제는 다양한 머신 러닝 방식에 네거티브 마커를 적용하여 미생물을 분류하는 방법 및 장치를 제공하는 것이다.

[9] 본 개시에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제 해결 수단

[10] 본 개시의 일 양상에 따른 유사 종을 식별하는 방법은, 입력된 샘플에 대한 제 1

질량 정보를 추출하는 단계; 상기 제 1 질량 정보에 기초하여, 적어도 네거티브 마커에 기반한 머신 러닝 모델을 이용하여 상기 입력된 샘플을 분류하는 단계; 및 상기 분류 결과에 기초하여 상기 입력된 샘플에 대한 종을 식별하는 단계를 포함할 수 있다.

- [11] 본 개시의 추가적인 양상에 따른 유사 종을 식별하는 장치는, 입력된 샘플에 대한 제 1 질량 정보를 추출하는 질량 분석기; 및 상기 제 1 질량 정보에 기초하여, 적어도 네거티브 마커 데이터베이스에 저장된 네거티브 마커에 기반한 머신 러닝 모델을 이용하여 상기 입력된 샘플을 분류하는 분류기를 포함하고, 상기 장치는 상기 분류 결과에 기초하여 상기 입력된 샘플에 대한 종을 식별할 수 있다.
- [12] 본 개시의 다양한 양상에 있어서, 포지티브 마커 및 상기 네거티브 마커를 이용하여 상기 입력된 샘플을 분류할 수 있다.
- [13] 본 개시의 다양한 양상에 있어서, 상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 유사 종에 속한 샘플들의 각각에 대해서 미리 추출될 수 있다.
- [14] 본 개시의 다양한 양상에 있어서, 상기 포지티브 마커는 대립 종에 비하여 타겟 종에서 빈번하게 나타나는 질량 정보를 포함할 수 있다.
- [15] 본 개시의 다양한 양상에 있어서, 상기 네거티브 마커는 타겟 종에 비하여 대립 종에서 빈번하게 나타나는 질량 정보를 포함할 수 있다.
- [16] 본 개시의 다양한 양상에 있어서, 상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 유사 종에 속한 샘플들의 각각에 대한 질량 스펙트럼에 대해서 설정된 빈에 기초하여 추출될 수 있다.
- [17] 본 개시의 다양한 양상에 있어서, 상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 질량 스펙트럼의 피크 값이 위치하는 빈의 번호의 집합으로 표현될 수 있다.
- [18] 본 개시의 다양한 양상에 있어서, 하나의 빈은 하나 이상의 다른 빈과 일부 중첩될 수 있다.
- [19] 본 개시의 다양한 양상에 있어서, 상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 질량 스펙트럼의 피크 값이 위치하는 빈의 빈도 정보에 기초하여 계산될 수 있다.
- [20] 본 개시의 다양한 양상에 있어서, 상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 빈 빈도 정보에 대해서 TF-IDF(Term Frequency-Inverse Document Frequency) 계산에 기초하여 추출될 수 있다.
- [21] 본 개시의 다양한 양상에 있어서, 상기 포지티브 마커는, 수학식

$$\text{TF-IDF}_{bin(i)} = \frac{F_{bin(i),sample_t}}{N_t} \times \log\left(\frac{N_o}{F_{bin(i),sample_o}}\right)$$

에 기초하여 계산되며, 상기 수학식에서 t는 타겟 종을 의미하고, o는 대립 종을 의미하고, N_t는 타겟 종에 대한 전체 개수를 의미하고, N_o는 대립 종에 대한 전체 개수를 의미하며,

$F_{bin(i)}$ 는 i -번째 빈에 대한 카운트 값을 의미할 수 있다.

[22] 본 개시의 다양한 양상에 있어서, 상기 포지티브 마커는, 상기 수학적식에 의해 계산된 TF-IDF 값이 소정의 임계치를 초과하는 경우에 상기 포지티브 마커로서 설정될 수 있다.

[23] 본 개시의 다양한 양상에 있어서, 상기 네거티브 마커는, 수학적식에 기초하여 계산되며, 상기

$$TF - IDF_{bin(i)} = \frac{F_{bin(i),sample_o}}{N_o} \times \log\left(\frac{N_t}{F_{bin(i),sample_t}}\right)$$

수학적식에서 t 는 타겟 종을 의미하고, o 는 대립 종을 의미하고, N_t 는 타겟 종에 대한 전체 개수를 의미하고, N_o 는 대립 종에 대한 전체 개수를 의미하며, $F_{bin(i)}$ 는 i -번째 빈에 대한 카운트 값을 의미할 수 있다.

[24] 본 개시의 다양한 양상에 있어서, 상기 네거티브 마커는, 상기 수학적식에 의해 계산된 TF-IDF 값이 소정의 임계치를 초과하는 경우에 상기 네거티브 마커로서 설정될 수 있다.

[25] 본 개시의 다양한 양상에 있어서, 상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 머신 러닝 모델의 학습을 위한 특징 추출을 위한 전처리 단계로서 생성될 수 있다.

[26] 본 개시의 다양한 양상에 있어서, 상기 제 1 질량 정보와, 하나 이상의 샘플의 각각에 대해서 미리 저장된 제 2 질량 정보에 기초하여 CCI(Composite Correlation Index)를 계산하고, 계산된 상기 CCI에 기초하여 상기 분류에 대한 후보를 결정하는 것을 더 포함할 수 있다.

[27] 본 개시에 대하여 위에서 간략하게 요약된 특징들은 후술하는 본 개시의 상세한 설명의 예시적인 양상일 뿐이며, 본 개시의 범위를 제한하는 것은 아니다.

발명의 효과

[28] 본 개시에 따르면, 질량 스펙트럼 분석에 관련된 네거티브 마커를 이용함으로써, 유사 종들간의 식별 성능을 개선하는 방법 및 장치가 제공될 수 있다.

[29] 본 개시에 따르면, 네거티브 마커를 이용함으로써 머신 러닝 기법에 무관하게 미생물 식별 성능을 개선하는 방법 및 장치가 제공될 수 있다.

[30] 본 개시에 따르면, 특징을 추출하기 위한 전처리를 적용함으로써, 머신 러닝 방식의 미생물 식별 성능을 개선하는 방법 및 장치가 제공될 수 있다.

[31] 본 개시에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[32] 도 1은 본 개시에 따른 마커 추출 과정을 설명하기 위한 도면이다.

- [33] 도 2는 본 개시에 따른 마커 추출에 이용되는 빈 방식에 대해서 설명하기 위한 도면이다.
- [34] 도 3은 본 개시에 따른 포지티브 마커 DB 및 네거티브 마커 DB에 저장되는 데이터의 예시들을 나타내는 도면이다.
- [35] 도 4는 본 개시에 따른 유사 종 분류를 위한 프로세스 프레임워크를 나타내는 도면이다.
- [36] 도 5는 본 개시에 따른 유사 종 분류를 위한 머신 러닝 모델을 설명하기 위한 도면이다.
- [37] 도 6은 본 개시에 따른 유사 종에 대한 컨퓨전 매트릭스 계산을 위한 머신 러닝 프로세스를 설명하기 위한 도면이다.
- [38] 도 7 및 도 8은 본 개시에 따른 마커-기반 식별 결과에 대한 평가 메트릭의 결과를 예시적으로 나타내는 도면이다.
- [39] 도 9는 본 개시에 따른 유사 종 식별 방법을 설명하기 위한 도면이다.

발명의 실시를 위한 최선의 형태

- [40] 이하에서는 첨부한 도면을 참고로 하여 본 개시의 실시예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나, 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다.
- [41] 본 개시의 실시예를 설명함에 있어서 공지 구성 또는 기능에 대한 구체적인 설명이 본 개시의 요지를 흐릴 수 있다고 판단되는 경우에는 그에 대한 상세한 설명은 생략한다. 그리고, 도면에서 본 개시에 대한 설명과 관계없는 부분은 생략하였으며, 유사한 부분에 대해서는 유사한 도면 부호를 붙인다.
- [42] 본 개시에 있어서, 어떤 구성요소가 다른 구성요소와 "연결", "결합" 또는 "접속"되어 있다고 할 때, 이는 직접적인 연결관계뿐만 아니라, 그 중간에 또 다른 구성요소가 존재하는 간접적인 연결관계도 포함할 수 있다. 또한 어떤 구성요소가 다른 구성요소를 "포함한다" 또는 "가진다"고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 배제하는 것이 아니라 또 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [43] 본 개시에 있어서, 제1, 제2 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용되며, 특별히 언급되지 않는 한 구성요소들간의 순서 또는 중요도 등을 한정하지 않는다. 따라서, 본 개시의 범위 내에서 일 실시예에서의 제1 구성요소는 다른 실시예에서 제2 구성요소라고 칭할 수도 있고, 마찬가지로 일 실시예에서의 제2 구성요소를 다른 실시예에서 제1 구성요소라고 칭할 수도 있다.
- [44] 본 개시에 있어서, 서로 구별되는 구성요소들은 각각의 특징을 명확하게 설명하기 위함이며, 구성요소들이 반드시 분리되는 것을 의미하지는 않는다. 즉, 복수의 구성요소가 통합되어 하나의 하드웨어 또는 소프트웨어 단위로

이루어질 수도 있고, 하나의 구성요소가 분산되어 복수의 하드웨어 또는 소프트웨어 단위로 이루어질 수도 있다. 따라서, 별도로 언급하지 않더라도 이와 같이 통합된 또는 분산된 실시예도 본 개시의 범위에 포함된다.

- [45] 본 개시에 있어서, 다양한 실시예에서 설명하는 구성요소들이 반드시 필수적인 구성요소들을 의미하는 것은 아니며, 일부는 선택적인 구성요소일 수 있다. 따라서, 일 실시예에서 설명하는 구성요소들의 부분집합으로 구성되는 실시예도 본 개시의 범위에 포함된다. 또한, 다양한 실시예에서 설명하는 구성요소들에 추가적으로 다른 구성요소를 포함하는 실시예도 본 개시의 범위에 포함된다.
- [46] 본 개시에서 사용하는 용어에 대한 정의는 다음과 같다.
- [47] 마커(marker): 타겟을 고유하게 식별하기 위해서 사용되는 특징
- [48] 포지티브 마커(positive marker): 대립 종(opposition species)에 비하여 타겟 종에서 보다 빈번하게 나타나는 특징
- [49] 네거티브 마커(negative marker): 타겟 종에 비하여 대립 종에서 보다 빈번하게 나타나는 특징
- [50] 빈(bin): 스펙트럼의 특정 구간
- [51] 본 개시에서 사용하는 약어에 대한 정의는 다음과 같다.
- [52] MALDI-TOF: Matrix-Assisted Laser Desorption/Ionization-Time-Of-Flight
- [53] MS: Mass Spectrometry
- [54] CCI: Composite Correlation Index
- [55] TF-IDF: Term Frequency-Inverse Document Frequency
- [56] 이하에서는 본 개시에 따른 네거티브 마커를 이용한 유사 종 식별 방법 및 장치에 대해서 설명한다.
- [57] MALDI-TOF MS는 단백질의 질량 조성(composition)을 통해서 미생물(microorganism)을 고속으로 식별(identify)할 수 있기 때문에 널리 사용되고 있다. 임의의 미생물에 대해서 추출된(extracted) 질량 조성 정보에 기초하여 해당 미생물을 다른 종(species)과 구별(distinguish)하는 마커를 선택함으로써 미생물이 식별될 수 있다. MALDI-TOF MS 등의 방식에 의해서 추출된 질량 정보와 머신 러닝(machine learning) 기법을 조합함으로써, 미생물 분류(classification)의 성능을 개선할 수 있다.
- [58] 미생물의 분류는 특히 마이코박테리아(mycobacteria)의 경우에 용이하지 않으면서도 매우 중요하다. 이는, 어떤 미생물 종들은 유사한 질량 조성을 보이지만, 서로 다른 병원균(pathogens)은 서로 다른 항생제(antibiotics)로 치료되어야 하기 때문이다. 유사한 미생물 종들의 MALDI-TOF 질량 스펙트럼 분석 패턴들이 서로 매우 유사하기 때문에, 종래의 방법을 통해서 유사한 미생물 종들을 정확하게 식별하는 것이 어렵다. 예를 들어, 결핵균(mycobacterium tuberculosis)의 경우에, 종들간의 질량 스펙트럼 패턴은 서로 매우 유사하고, 다른 박테리아에 비해서 식별의 정확도가 상대적으로 낮다.

- 각각의 미생물 종의 구성요소들이 서로 매우 유사하지만, 환자에 대한 처방은 각각의 종에 대해서 서로 달라야만 하기 때문에, 미생물 유사 종에 대한 분류는 매우 중요하다. 또한, CCI는 질량 스펙트럼 분석에 기초하여 유사한 박테리아를 찾는 효율적인 방법이지만, 마이코박테리움 앱시서스(*mycobacterium abscessus*) 그룹과 같은 유사한 종들을 정확하게 분류할 수는 없다. 따라서, 종래의 방식과는 다른 새로운 방식으로 미생물을 식별 또는 분류하는 방안이 요구된다.
- [59] 본 개시에 따르면, 네거티브 마커를 이용함으로써 미생물 식별 성능을 높일 수 있다. 또한, 본 개시에 따르면, 포지티브 마커 및 네거티브 마커를 이용한 새로운 머신 러닝 기법을 적용함으로써, 미생물 질량 스펙트럼 분석에 있어서의 식별 및 분류 성능을 높일 수 있다. 또한, 본 개시에 따르면 새로운 머신 러닝에 이용되는 특징(feature)에 대한 전처리를 적용하는 새로운 방식을 제공한다. 예를 들어, 특징에 대한 전처리는 네거티브 마커 추출을 포함한다. 또한, 특징에 대한 전처리는 포지티브 마커 및 네거티브 마커를 개별적으로(*separately*) 추출하는 것을 포함한다. 이에 따라, 어떠한 머신 러닝 기법을 적용하더라도 유사한 종들의 식별 성능을 높일 수 있다. 즉, 머신 러닝 기법에 무관하게, 미생물의 식별 및 분류의 성능을 높일 수 있다.
- [60] 본 개시에서는 마이코박테리움(*mycobacterium*) 앱시서스(*abscessus*) 그룹과 마이코박테리움 포튜이툼(*fortuitum*) 그룹에 대한 아류(subtype) 또는 아종(subspecies)의 식별 또는 분류를 대표적인 예시로 들어 설명한다. 그러나, 본 개시의 범위가 이에 제한되는 것은 아니며, 다양한 미생물의 유사 종에 대해서 네거티브 마커를 이용한 식별 또는 분류 방안이 적용되는 것을 포함한다.
- [61] 또한, 본 개시에서는 머신 러닝 기법으로서 SVM(support vector machine)을 대표적인 예시로 들어 설명한다. 그러나, 본 개시의 범위가 이에 제한되는 것은 아니며, k-NN(k-nearest neighbor), 뉴럴 네트워크(neural network), 랜덤 포레스트 알고리즘(random forest algorithm) 등의 다양한 머신 러닝 기법에, 본 개시에 따른 네거티브 마커를 이용한 유사 종 식별 또는 분류 방안이 적용되는 것을 포함한다.
- [62] 이하에서는, 포지티브 마커 및 네거티브 마커 추출에 대해서 먼저 설명하고, 추출된 마커를 이용하여 유사 종들을 분류하는 모델에 대해서 설명한다.
- [63] 도 1은 본 개시에 따른 마커 추출 과정을 설명하기 위한 도면이다.
- [64] 본 개시에서는 마이코박테리아의 각각의 아류(subtype)로부터 포지티브 마커 및 네거티브 마커를 추출하여, 머신 러닝 모델로 사용하는 새로운 프레임워크를 포함한다. 이와 같이 포지티브 및 네거티브 마커를 사용함으로써, 본 개시에 따른 모델은 어떠한 타입의 머신 러닝에 있어서도 아종(subspecies) 분류의 정확성을 크게 높일 수 있다.
- [65] 도 1에서 질량 정보 DB(110)는 하나 이상의 미생물 그룹에 속한 종들에 대한 질량 정보의 데이터셋(dataset)을 포함할 수 있다. 구체적으로, 질량 정보 DB(110)는 하나 이상의 미생물 그룹의 각각에 속한 하나 이상의 종의 각각에

대한 질량 정보를 포함할 수 있다. 예를 들어, 질량 정보는 각각의 미생물 샘플에 대한 MALDI-TOF MS 분석에 의해서 얻어질 수 있다.

[66] 표 1은 질량 정보 DB(110)에 포함되는 데이터셋에 대한 통계의 일례를 나타낸다.

[67] [표1]

Group	Species	Number of spectra
M. abscessus	M. abscessus	167
	M. bolletii	95
	M. massiliense	163
M. fortuitum	M. fortuitum	124
	M. conceptionense	109
	M. neworleansense	18
	M. peregrinum	58
	M. porcinum	62

[68] 표 1에서는 M. abscessus 그룹에는 M. abscessus, M. bolletii 및 M. massiliense가 속하고, 각각의 샘플에 대한 질량 스펙트럼의 개수는 167, 95 및 163개임을 나타낸다. 또한, M. fortuitum 그룹에는 M. fortuitum, M. conceptionense, M. neworleansense, M. peregrinum 및 M. porcinum가 속할 수 있고, 각각의 샘플에 대한 질량 스펙트럼의 개수는 124, 109, 18, 58 및 62 개임을 나타낸다. 질량 정보 DB(110)에는 각각의 종에 대한 실제 질량 스펙트럼 정보가 포함되는 것으로 가정한다.

[69] 도 1의 타겟 종(target species)에 대한 마커 추출(marker extraction) 과정(120)에서는, 질량 정보 DB(110)에 포함된 데이터 중에서 특정 타겟 종에 대한 질량 정보에 기초하여 마커를 추출할 수 있다. 예를 들어, 포지티브 마커는 다른 유사 종(예를 들어, 대립 종)에 비하여 타겟 종에서 빈번하게(frequently) 나타나는 질량 정보를 포함할 수 있다. 마커 추출(120)의 결과는 포지티브 마커 DB(130)에 저장 및 유지될 수 있다.

[70] 도 1의 대립 종(opposition species)에 대한 마커 추출(marker extraction) 과정(140)에서는, 질량 정보 DB(110)에 포함된 데이터 중에서 특정 대립 종에 대한 질량 정보에 기초하여 마커를 추출할 수 있다. 예를 들어, 네거티브 마커는 타겟 종에 비하여 대립 종에서 빈번하게 나타나는 질량 정보를 포함할 수 있다. 마커 추출(140)의 결과는 네거티브 마커 DB(150)에 저장 및 유지될 수 있다.

[71] 예를 들어, M. abscessus, M. bolletii 및 M. massiliense는 유사한 그룹이다. 만약 선택된 타겟이 M. abscessus인 경우, M. bolletii 및 M. massiliense는 대립 종이 될 수 있다.

[72] 이와 같이, 마이코박테리움 데이터셋으로부터 특정 박테리아의 특징(feature)를 나타내는 마커를 추출할 수 있다. 마커 추출의 일례로서 TF-IDF 방식이 적용될 수 있으며, 이에 대해서는 후술한다.

- [73] 도 2는 본 개시에 따른 마커 추출에 이용되는 빈 방식에 대해서 설명하기 위한 도면이다.
- [74] MALDI-TOF MS는, 동일한 실험이 반복되더라도 반드시 동일한 결과를 내지는 않는다. 동일한 분자(molecule)에 대해서도, 전체 비행 시간(total flight time)은 이온 비행의 각도에 따라서 약간씩 상이할 수 있다. 이는 질량 스펙트럼의 피크 시프트(peak shift)의 원인이 될 수 있다.
- [75] 이러한 피크 시프트를 고려하기 위해서, 본 개시에서는 도 2에서 나타내는 바와 같이 질량에 대한 빈(bin)을 적용하고 빈 윈도우들을 일부 중첩(overlap)할 수 있다.
- [76] 어떤 샘플에 대한 질량 스펙트럼에서 피크 값이 위치하는 빈(bin)의 빈도(frequency)를 계산함으로써, 해당 샘플의 질량 스펙트럼에 대한 특성을 빈(bin) 번호(number)의 집합(aggregation)으로 표현할 수 있다. 이에 따라, 특정 샘플에 대한 특징 값을 보다 정확하게 추출할 수 있다.
- [77] 이와 같이 본 개시에 따르면 질량 정보에 빈을 적용하는 데이터 전처리(preprocessing)를 적용한다. 이에 따라, 피크 시프트와 같은 관측 에러(observation errors)의 영향을 줄일 수 있다.
- [78] 구체적으로, 포지티브 마커 DB(130) 및 네거티브 마커 DB(150)의 각각에 저장되는 질량 정보는, 질량 빈 번호의 집합으로 구성될 수 있다. 하나의 질량 빈은 질량 스펙트럼에서 특정 구간에 해당할 수 있다. 또한, 하나의 질량 빈은 하나 이상의 다른 질량 빈과 일부 중첩(overlap)될 수 있다.
- [79] 예를 들어, 질량 스펙트럼의 전체 범위를 동일한 크기의 100개의 빈이 커버하는 경우를 가정한다. 빈 번호는 낮은 스펙트럼 구간부터 순서대로 bin1, bin2, bin3, ..., bin100으로 할당될 수 있다. 도 2의 예시에서와 같이 bin29의 높은 질량 값 구간의 일부는 bin30의 낮은 질량 값 구간의 일부와 중첩될 수 있다. 또한, bin30의 낮은 질량 값 구간의 일부는 bin29의 높은 질량 값 구간의 일부와 중첩될 수 있고, bin30의 높은 질량 값 구간의 일부는 bin31의 낮은 질량 값 구간의 일부와 중첩될 수 있다. 그러나, 본 개시의 범위가 전술한 예시로 제한되는 것은 아니며, 어떤 질량 값 구간이 셋 이상의 빈이 중첩되는 구간으로 설정될 수도 있고, 어떤 질량 값 구간은 하나의 빈에 의해서만 커버될 수도 있다.
- [80] 도 2의 예시에서 특정 샘플의 질량 스펙트럼의 일부 구간에서, 질량대전하비(m/z)의 신호 강도에서 두 개의 피크(210, 220)가 검출되는 것을 나타낸다. 검출된 피크(210)는 bin29에 해당하는 것으로 확인되는 이벤트(check1)가 발생하고, 또 다른 검출된 피크(220)는 bin30에도 해당하며 bin31에도 해당하는 것으로 확인되는 이벤트(check2)가 발생할 수 있다. 이에 따라, bin29의 빈도(frequency)는 check1 이벤트로 인하여 +1 카운트되고, bin30의 빈도는 check2 이벤트로 인하여 +1 카운트되고, bin31의 빈도는 check2 이벤트로 인하여 +1 카운트될 수 있다. 그리고, bin32에 해당하는 구간에서는 피크 값이 검출되지 않으므로, bin32의 빈도는 0으로 카운트된다.

- [81] 이와 같이, 빈이라고 칭하는 소정의 인터벌에 오리지널 데이터 값이 속하는 경우, 해당 데이터 값은 상기 소정의 인터벌의 대표 값으로 치환될 수 있다. 인터벌의 대표 값은 일반적으로 해당 인터벌의 중앙 값(central value)일 수 있지만, 이에 제한되는 것은 아니며 해당 인터벌에 속하는 시작 값, 끝 값 또는 임의의 값을 대표 값으로 정의할 수도 있다. 예를 들어, 도 2의 예시에서 bin29의 대표 값은 해당 빈의 번호, 즉, 29로 주어질 수 있다.
- [82] 빈의 크기가 큰 경우(즉, 전체 스펙트럼 구간을 커버하는 빈의 개수가 적은 경우)에 샘플을 다른 유사 종과 정확하게 구분하는 성능이 낮아질 수 있다. 반대로 빈의 크기가 좁은 경우(즉, 전체 스펙트럼 구간을 커버하는 빈의 개수가 많은 경우)에는 관측 에러(예를 들어, 피크 시프트)의 영향을 줄이기가 어려워질 수 있다. 이러한 점을 고려하여, 본 개시에서 예시적인 적절한 빈의 크기는 $20m/z$ 로 설정할 수 있다.
- [83] 또한, 빈 윈도우가 중첩되는 범위는 도 2의 예시와 같이 매 짝수 번째 빈의 시작 위치와 종료 위치가 중첩되지 않지만 연속적이고, 매 홀수 번째 빈의 시작 위치와 종료 위치가 중첩되지 않지만 연속적인 것으로 설정한다. 예를 들어, 도 2에서 bin29의 종료 위치는 bin31의 시작 위치와 중첩되지 않지만 연속적인 값을 커버하도록 설정될 수 있다.
- [84] 본 개시의 범위는 위와 같은 예시적인 빈의 크기 및 중첩 범위에 의해서 제한되지 않으며, 데이터셋의 특성을 고려하여 적절하게 설정할 수 있다. 즉, 본 개시의 특징은 설정된 빈을 이용하여 포지티브 마커 및 네거티브 마커를 추출하기 위한 전처리를 적용하는 것 자체에 있으며, 빈의 크기, 개수, 중첩 범위 등의 구체적인 값으로 제한되는 것은 아니다.
- [85] 도 3은 본 개시에 따른 포지티브 마커 DB 및 네거티브 마커 DB에 저장되는 데이터의 예시들을 나타내는 도면이다.
- [86] 도 2와 같이 샘플에 대한 질량 데이터 특징이 빈 번호의 집합 형태로 DB에 저장되면, 해당 정보로부터 포지티브 마커 및 네거티브 마커를 추출할 수 있다. 즉, 빈 빈도(bin frequency)를 계산함으로써, 타겟 종 또는 대립 종에서 어떤 빈(들)이 빈번하게 나타나는지를 검출할 수 있다. 이와 같은 각각의 종에 대한 빈 빈도 정보에 대해서 조정된(adjusted) TF-IDF를 계산함으로써, 최종적으로 포지티브 마커 및 네거티브 마커를 추출할 수 있다. 예를 들어, 도 1에서 타겟 종에 대한 마커 추출(120) 및 대립 종에 대한 마커 추출(130)에 있어서 후술하는 TF-IDF 계산이 적용될 수 있다.
- [87] 수학적 식 1은 포지티브 마커를 추출하기 위한 수학적 식을 나타낸다.
- [88] [수식1]

$$TF - IDF_{bin(i)} = \frac{F_{bin(i),sample_t}}{N_t} \times \log\left(\frac{N_o}{F_{bin(i),sample_o}}\right)$$

- [89] 수학적 식 1에서 t는 타겟 종을 의미하고, o는 대립 종을 의미한다. Nt는 타겟 종에

대한 전체 개수를 의미하고, N_o 는 대립 종에 대한 전체 개수를 의미한다.
 $F_{bin(i)}$ 는 i -번째 빈에 대한 카운트 값을 의미한다.

[90] 또한, TF-IDF 임계치(threshold)를 포지티브 마커와 네거티브 마커를 구분하는 기준으로 삼을 수 있다. 예를 들어, 타겟 종에서의 빈 빈도가 85%이고, 대립 종에서의 빈 빈도가 15%인 경우, TF-IDF 임계치는 0.676498일 수 있다. 따라서, 각각의 빈에서의 TF-IDF 값이 임계치(예를 들어, 0.676498)를 초과하는 경우에 해당 빈을 포지티브 마커로 설정할 수 있다.

[91] 수학식 2는 네거티브 마커를 추출하기 위한 수학적식을 나타낸다.

[92] [수식2]

$$TF - IDF_{bin(i)} = \frac{F_{bin(i),sample_o}}{N_o} \times \log\left(\frac{N_t}{F_{bin(i),sample_t}}\right)$$

[93] 수학식 2는 수학식 1에서 타겟 종을 대립 종으로 교환한 것에 해당한다. 즉, 수학식 2에서 t 는 타겟 종을 의미하고, o 는 대립 종을 의미한다. N_t 는 타겟 종에 대한 전체 개수를 의미하고, N_o 는 대립 종에 대한 전체 개수를 의미한다. $F_{bin(i)}$ 는 i -번째 빈에 대한 카운트 값을 의미한다. 수학식 2와 같이 계산된 TF-IDF 결과에 대한 랭킹 및 스케일에 기초하여 의미 있는 마커를 식별할 수 있다.

[94] 또한, TF-IDF 임계치(threshold)를 포지티브 마커와 네거티브 마커를 구분하는 기준으로 삼을 수 있다. 예를 들어, 대립 종에서의 빈 빈도가 85%이고, 타겟 종에서의 빈 빈도가 15%인 경우, TF-IDF 임계치는 0.676498일 수 있다. 따라서, 각각의 빈에서의 TF-IDF 값이 임계치(예를 들어, 0.676498)를 초과하는 경우에 해당 빈을 네거티브 마커로 설정할 수 있다.

[95] 수학식 1 및 2와 같이 계산된 TF-IDF 결과에 대한 랭킹 및 스케일에 기초하여 1의 의미 있는 마커를 식별할 수 있다. 이를 이용하여, 각각의 박테리아에 대한 포지티브 마커 DB 및 네거티브 마커 DB를 도 3과 같이 구성할 수 있다.

[96] 예를 들어, 도 3에서 박테리아 식별자(Bacteria_ID)가 a_1 인 박테리아에 대한 포지티브 마커는 $bin_1, bin_{31}, bin_{42}, \dots$ 이라는 빈 번호 집합(즉, 피크가 검출된 빈 번호의 집합)에 대한 정보가 저장될 수 있다. 또한, 동일한 a_1 식별자를 가지는 박테리아에 대한 네거티브 마커는 $bin_7, bin_{35}, bin_{49}, \dots$ 이라는 빈 번호 집합에 대한 정보가 저장될 수 있다. 또한, 포지티브 마커 및 네거티브 마커는 각각의 박테리아(예를 들어, a_2, a_3, a_4, \dots)에 대해서 저장될 수 있다.

[97] 이와 같이 데이터셋에 대한 전처리 결과로 포지티브 마커 및 네거티브 마커가 결정될 수 있고, 이러한 전처리 결과를 이용하여(특히, 네거티브 마커를 이용하여) 미지의 샘플에 대한 질량 특성을 분석함으로써 해당 샘플이 어떤 박테리아에 해당하는지를 정확하게 식별 또는 분류할 수 있다.

[98] 이하에서는, 전술한 포지티브 마커 및 네거티브 마커 추출에 대한 설명에 이어서, 추출된 마커를 이용하여 유사 종들을 분류하는 모델에 대해서 설명한다.

- [99] 도 4는 본 개시에 따른 유사 종 분류를 위한 프로세스 프레임워크를 나타내는 도면이다.
- [100] 새로운 샘플(410)이 유사 종 분류 프로세스에 입력되면, 질량 분석기(420)에서 해당 샘플에 대한 질량 분석이 수행될 수 있다. 질량 분석의 결과로 해당 샘플에 대한 질량 패턴(425)이 추출될 수 있다. 예를 들어, MALDI-TOF 방식으로 샘플에 대한 질량 스펙트럼 분석이 수행될 수 있으며, 질량 패턴은 질량 스펙트럼의 형태로 획득될 수 있다. 즉, 질량 정보는 질량 및 강도(intensity) 값을 포함할 수 있다.
- [101] 유사도 계산기(430)에서는, 샘플에 대해서 추출된 질량 패턴(425) 정보와 데이터베이스(436)에 저장된 정보와의 유사도(similarity)를 계산할 수 있다. 예를 들어, 유사도의 계산은 입력된 샘플에 대해서 추출된 질량 패턴(425) 정보 및 데이터베이스(436)에 저장된 정보에 대한 CCI를 계산함으로써 수행될 수 있다. 구체적으로, 입력된 샘플(410)에 대해서 획득된 질량 및 강도 값과, 데이터베이스(436)에 저장된 샘플들에 대해서 미리 획득되어 있는 질량 및 강도 값에 대해서, CCI 계산을 이용하여 양자간의 유사도를 획득할 수 있다.
- [102] CCI 계산을 통해서 유사한 그룹을 추출할 수 있지만, 유사한 그룹들 중에서 타겟을 정확하게 식별하기에는 불충분하다. 이러한 문제를 해결하기 위해서, 머신 러닝 모델이 본 개시에 따른 네거티브 마커를 이용하여 분류를 학습하도록 함으로써, CCI 계산 결과에서 유사 종을 정확하게 분류할 수 있다. 보다 구체적으로, 본 개시에 따르면 포지티브 마커 및 네거티브 마커를 이용하여 머신 러닝 모델이 분류를 학습하도록 함으로써, CCI 계산 결과로부터 유사 종을 보다 정확하게 분류할 수 있다.
- [103] 예를 들어, CCI 비교기(432)에서는 입력된 샘플(410)에 대해서 추출된 질량 정보(즉, 제 1 질량 정보)와, 데이터베이스(436)에 미리 저장되어 있는 샘플에 대한 질량 정보(즉, 제 2 질량 정보)에 기초하여, CCI를 계산할 수 있다. 데이터베이스(436)에는 하나 이상의 샘플에 대한 질량 정보가 미리 저장되어 있을 수도 있으므로, CCI 계산은 데이터베이스(436)의 하나 이상의 샘플의 각각에 대한 제 2 질량 정보에 기초하여 수행될 수 있다. 즉, 제 1 질량 정보와 하나 이상의 제 2 질량 정보의 각각에 대해서 CCI 계산이 수행될 수 있다.
- [104] CCI 비교기(432)는 제 1 질량 정보와 하나 이상의 제 2 질량 정보 각각에 대한 CCI 값을 계산함으로써, 입력된 샘플(410)에 매칭되는 데이터베이스(436)에 저장된 샘플의 후보를 결정할 수 있다. 이와 같이 CCI 계산을 통해서 압축된 후보(434)를 지시하는 정보는 분류기(440)로 전달될 수 있다.
- [105] 분류기(440)에서는 CCI 계산을 통해서 압축된 후보(434)에 대한 머신 러닝 모델을 이용한 분류 프로세스를 수행할 수 있다. 분류기(440)는 모델 분류자(450) 및 학습 모델(460)을 포함할 수 있다. 학습 모델(460)은 포지티브 마커 DB(470)에 저장된 정보 및 네거티브 마커 DB(480)에 저장된 정보를 특징 값(feature value)으로 사용하여 각각의 종에 대한 분류를 학습(465)할 수 있다. 모델

분류자(455)에서는 학습 모델(460)에 기초하여 새로운 샘플(410)에 대한 유사 종 분류(455)를 수행하고, 그 결과로 특정 클래스가 유도(derive)될 수 있다. 유도된 결과는 다시 머신 러닝의 샘플로 사용될 수 있다.

- [106] 이와 같이, 새로운 샘플에 대한 데이터 값이 머신 러닝 분류자(classifier)에 입력되면, 미리 학습된 모델에 기초하여 특정 클래스가 유도될 수 있다. 또한, 분류 결과에 기초하여, 입력된 새로운 샘플에 대한 종을 식별할 수 있다.
- [107] 도 5는 본 개시에 따른 유사 종 분류를 위한 머신 러닝 모델을 설명하기 위한 도면이다.
- [108] 도 5의 예시에서는 포지티브 마커 및 네거티브 마커를 특징(feature)으로 사용하는 머신 러닝 프로세스의 일례를 나타낸다.
- [109] 전술한 바와 같이, 포지티브 마커는 타겟 종에 대한 질량 정보를 포함할 수 있고, 네거티브 마커는 대립 종에 대한 질량 정보를 포함할 수 있다. 각각의 샘플에 대해서 질량 빈 정보가 평가(evaluate)될 수 있다. 예를 들어, 질량 빈 정보의 평가는 부울 연산자(Boolean operator)를 이용하여 수행될 수 있다.
- [110] 도 5의 예시에서, 샘플 1(sample 1)에 대한 포지티브 마커 체크(check) 결과는 111101으로 표시되고, 네거티브 마커 체크 결과는 000000으로 표시된다. 여기서 1은 참(true)을 의미하고, 0은 거짓(false)을 의미한다. 이에 따라, 샘플 1을 클래스 1(class 1)로 분류하는 것으로 학습될 수 있다. 이와 마찬가지로, 포지티브 마커 체크 결과가 네거티브 마커 체크 결과에 비하여 상대적으로 더 매칭되는 체크 결과를 포함하는 샘플 2 내지 4의 경우에, 해당 샘플은 클래스 1로 분류하는 것으로 학습될 수 있다. 한편, 샘플 40 내지 42의 경우에는, 포지티브 마커 체크 결과가 네거티브 마커 체크 결과에 비하여 상대적으로 덜 매칭되는 체크 결과를 포함하므로, 해당 샘플들은 클래스 2로 분류하는 것으로 학습될 수 있다.
- [111] 도 5의 예시에서 나타내는 바와 같이 타겟 종과 대립 종 간에는 명확한 차이점이 존재한다. 이와 같이, 포지티브 마커 및 네거티브 마커를 이용함으로써 머신 러닝 모델에 기초한 분류자의 성능이 크게 향상될 수 있다.
- [112] 도 6은 본 개시에 따른 유사 종에 대한 컨퓨전 매트릭스 계산을 위한 머신 러닝 프로세스를 설명하기 위한 도면이다.
- [113] 도 6의 예시에서 샘플 1에 대해서 종 A(species A)의 마커 1(marker 1), 종 A의 마커 2, ... 종 A의 마커 35, 종 B의 마커 1,..., 종 B의 마커 45에 대한 체크 결과가 11...01...0로 표시된다. 다음으로, 샘플 2 내지 샘플 95의 각각에 대해서도, 종 A의 마커 1 내지 종 B의 마커 45에 대한 체크 결과가 예시적으로 표시된다. 이러한 마커 체크 결과에 기초하여 머신 러닝에서 각각의 샘플을 클래스 1, 클래스 2, ... 등으로 분류할 수 있고 이러한 분류 결과가 학습될 수 있다.
- [114] 또한, 도 6의 예시에서 종 A의 마커 1에 대해서 샘플 1 내지 95의 체크 결과가 11111...00000으로 표시된다. 또한, 도 6에서는 종 A의 마커 2 내지 종 B의 마커 45의 각각에 대해서, 샘플 1 내지 95의 체크 결과가 예시적으로 표시된다.
- [115] 이와 같이, 도 6의 예시에서 종(species)은 포지티브 마커와 네거티브

마커로부터 부울 벡터(Boolean vector)를 가진다. 이러한 벡터들은 컨퓨전 매트릭스(confusion matrix) 계산을 위해서 머신 러닝 모델에 사용될 수 있다.

[116] 유사 종에 대한 전체 마커를 학습함으로써, 머신 러닝 모델에 기초하여 샘플을 보다 정확하게 분류할 수 있다. 컨퓨전 매트릭스를 계산함으로써, 서로 다른 그룹(예를 들어, 서로 다른 종)에 대해서 특정 엔트리를 식별할 수 있다. 또한, 컨퓨전 매트릭스를 계산함으로써, 본 개시에 따른 포지티브 마커 및 네거티브 마커를 이용한 모델에 대한 기준 에러(standard error)를 확인하고, 정확하게 종을 식별한 비율(즉, 퍼센티지)을 확인함으로써 해당 종에 대한 내부 안정성(internal stability)를 측정할 수 있다. 이러한 컨퓨전 매트릭스 계산은 SVM, k-NN, 뉴럴 네트워크, 랜덤 포레스트 등의 다양한 머신 러닝 기법에 대해서 확인될 수 있다.

[117] 평가 매트릭(evaluation metric)으로서, 두 가지 기법을 적용할 수 있다.

[118] 첫 번째는 프리시전(precision), 리콜(recall) 및 f-스코어(f-score)를 이용하는 기법이고, 두 번째는 정확도(accuracy)를 이용하는 기법이다.

[119] 프리시전, 리콜 및 f-스코어는 아래의 수학적식 3과 같이 정의된다.

[120] [수식3]

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$f\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

[121] 수학적식 3에서 tp는 진양성(true positive)을 의미하고, fp는 위양성(false positive)을 의미하며, fn은 위음성(false negative)을 의미한다. 또한, f-스코어는 프리시전과 리콜의 조화 평균(harmonic mean)에 해당한다.

[122] 다음으로, 정확도(accuracy)는 아래의 수학적식 4와 같이 정의된다.

[123] [수식4]

$$\text{Accuracy} = \frac{tp + tn}{tp + fp + fn + tn}$$

[124] 수학적식 4에서 tp는 진양성(true positive)을 의미하고, fp는 위양성(false positive)을 의미하고, tn은 진음성(true negative)을 의미하며, fn은 위음성(false negative)을 의미한다.

[125] 아래의 표 2 및 표 3은 표 1과 같은 테스트셋에 대한 유사 종 식별 결과를 포함하는 다중-클래스 컨퓨전 매트릭스를 나타낸다.

[126] [표2]

Marker-based SVM model (M. abscessus group)	P1	P2	P3
T1	94.05%	4.17%	1.79%
T2	15.42%	77.71%	6.88%
T3	0.73%	1.83%	97.44%

[127] 표 2는 *M. abscessus* 그룹에 대한 마커-기반 SVM 모델의 식별 결과를 나타낸다. T는 정확한 종을 의미하고, P는 예측된 종을 의미한다. 인덱스 1, 2 및 3은 각각 *M. abscessus*, *M. bolletii* 및 *M. massiliense*를 의미한다.

[128] [표3]

Marker-based SVM model (<i>M. fortuitum</i> group)	P1	P2	P3	P4	P5
T1	100%	0%	0%	0%	0%
T2	0%	100%	0%	0%	0%
T3	1.11%	0%	82.22%	15.56%	1.11%
T4	1.03%	0%	0%	98.62%	0.34%
T5	0.32%	0%	0%	1.94%	97.74%

[129] 표 3은 *M. fortuitum* 그룹에 대한 마커-기반 SVM 모델의 식별 결과를 나타낸다. T는 정확한 종을 의미하고, P는 예측된 종을 의미한다. 인덱스 1, 2, 3, 4 및 5는 각각 *M. fortuitum*, *M. conceptionense*, *M. neworleansense*, *M. peregrinum* 및 *M. porcinum*를 의미한다.

[130] 표 2 및 표 3은 모두 매우 정확한 종 구별 결과를 나타낸다. 표 2에서 *M. M. bolletii*를 예측하는 것이 다른 종을 예측하는 것에 비하여 어려운 것을 나타내고, 표 3에서는 T3가 패턴을 학습하기에 샘플이 부족한 것을 나타내지만, 샘플이 충분한 경우 분류 성능이 매우 높은 것을 나타낸다. 이러한 패턴은 아래의 표 4 내지 표 9와 같은 다른 학습 모델에 대해서도 관찰된다.

[131] 아래의 표 4, 6 및 8은 표 2와 같이 *M. abscessus* 그룹에 대한 마커-기반 머신 러닝 모델(각각 k-NN, 뉴럴 네트워크, 랜덤 포레스트 모델)의 식별 결과를 나타내고, 표 5, 7 및 9는 *M. fortuitum* 그룹에 대한 마커 기반 머신 러닝 모델(각각 k-NN, 뉴럴 네트워크, 랜덤 포레스트 모델)의 식별 결과를 나타낸다.

[132] [표4]

Marker-based k-NN model (<i>M. abscessus</i> group)	P1	P2	P3
T1	93.45%	5.00%	1.55%
T2	15.42%	69.38%	15.21%
T3	0.24%	0.73%	99.02%

[133] [표5]

Marker-based k-NN model (<i>M. fortuitum</i> group)	P1	P2	P3	P4	P5
T1	99.03%	0.97%	0%	0%	0%
T2	0%	100%	0%	0%	0%
T3	2.22%	0%	87.78%	7.78%	2.22%
T4	2.76%	0%	0%	95.86%	1.38%
T5	0%	0.32%	0%	0%	99.68%

[134] [표6]

Marker-based neural network model (M. abscessus group)	P1	P2	P3
T1	88.21%	10.60%	1.19%
T2	22.50%	64.79%	12.71%
T3	0.49%	10.98%	88.54%

[135] [표7]

Marker-based neural network model (M. fortuitum group)	P1	P2	P3	P4	P5
T1	93.51%	5.97%	0.32%	0%	0%
T2	4.18%	85.82%	8.91%	1.09%	0%
T3	0%	15.56%	46.67%	28.89%	8.89%
T4	0.69%	2.76%	21.38%	63.45%	11.72%
T5	0%	0%	0.65%	9.03%	90.32%

[136] [표8]

Marker-based random forest model (M. abscessus group)	P1	P2	P3
T1	92.38%	5.00%	2.62%
T2	13.96%	76.04%	10.00%
T3	1.22%	1.83%	96.95%

[137] [표9]

Marker-based random forest model (M. fortuitum group)	P1	P2	P3	P4	P5
T1	98.71%	1.29%	0%	0%	0%
T2	0%	100%	0%	0%	0%
T3	6.67%	0%	85.56%	6.67%	1.11%
T4	1.72%	0%	0%	98.28%	0%
T5	0%	0%	0%	0%	100%

[138] 도 7 및 도 8은 본 개시에 따른 마커-기반 식별 결과에 대한 평가 메트릭의 결과를 예시적으로 나타내는 도면이다.

[139] 도 7은 *M. abscessus* 그룹에 대해서 포지티브 마커 및 네거티브 마커를 모두 사용한 식별 결과와, 포지티브 마커만을 사용한 식별 결과에 대해서, 각각의 머신 러닝 기법 별로 정확도(accuracy) 및 f-스코어의 값을 나타낸다.

[140] 도 8은 *M. fortuitum* 그룹에 대해서 포지티브 마커 및 네거티브 마커를 모두 사용한 식별 결과와, 포지티브 마커만을 사용한 식별 결과에 대해서, 각각의 머신 러닝 기법 별로 정확도(accuracy) 및 f-스코어의 값을 나타낸다.

- [141] 도 7 및 도 8에서 나타내는 바와 같이, 포지티브 마커만을 사용하는 종래의 머신러닝 모델에 비하여, 본 개시에 따른 포지티브 마커 및 네거티브 마커를 사용하는 머신러닝 모델에 비하여, 대략 1 내지 5% 정도로 정확도가 향상되는 것을 알 수 있다. 따라서, 본 개시에 따른 네거티브 마커를 이용하는 유사 종 식별 방법은, 머신러닝 기법과 무관하게 유사 종 식별 성능을 향상시킬 수 있다.
- [142] 도 9는 본 개시에 따른 유사 종 식별 방법을 설명하기 위한 도면이다.
- [143] 단계 S910에서 입력된 샘플에 대한 제 1 질량 정보를 추출할 수 있다. 예를 들어, MALDI-TOF MS 방식에 기초하여 입력된 샘플에 대한 질량 스펙트럼 또는 질량 패턴 정보를 추출할 수 있다.
- [144] 단계 S920에서 단계 S910에서 추출된 제 1 질량 정보와, 하나 이상의 샘플의 각각에 대해서 미리 저장된 제 2 질량 정보에 기초하여 CCI를 계산할 수 있다. 제 2 질량 정보는 하나 이상의 샘플에 대해서 미리 추출되어 데이터베이스에 저장된 것일 수 있다.
- [145] 단계 S930에서 단계 S920의 CCI 계산 결과에 기초하여 분류에 대한 후보를 결정할 수 있다.
- [146] 상기 단계 S920 및 S930은 유사 종 분류의 후보를 결정하는 측면에서, 후속하는 마커-기반 머신러닝 모델을 이용한 유사 종 분류의 복잡도를 낮추고 성능을 높이는 데에 도움을 줄 수 있다. 본 개시의 범위는, 만약 단계 S920 및 S930이 수행되지 않는 경우라도, 제 1 질량 정보에 기초한 마커-기반 머신러닝 모델을 이용함으로써 입력된 샘플을 유사 종 간에 충분히 분류할 수 있다.
- [147] 단계 S940에서 단계 S910에서 추출된 제 1 질량 정보에 기초하여, 마커-기반 머신러닝 모델을 이용하여 입력된 샘플을 분류할 수 있다. 마커-기반 머신러닝 모델은 적어도 네거티브 마커를 이용한 머신러닝 모델을 포함할 수 있다. 또한, 마커-기반 머신러닝 모델은 포지티브 마커 및 네거티브 마커를 이용한 머신러닝 모델을 포함할 수 있다.
- [148] 포지티브 마커 및 상기 네거티브 마커의 각각은, 유사 종에 속한 샘플들의 각각에 대해서 미리 추출될 수 있다. 예를 들어, 포지티브 마커 및 상기 네거티브 마커의 각각은, 유사 종에 속한 샘플들의 각각에 대한 질량 스펙트럼에 대해서 설정된 빈에 기초하여 추출될 수 있다. 이와 같이 샘플들의 질량 정보에 대해서 빈을 적용하여 포지티브 마커 및 네거티브 마커를 추출하는 것은, 머신러닝 모델의 학습을 위한 특징을 추출하기 위한 전처리 과정으로서 수행될 수 있다.
- [149] 단계 S950에서 단계 S940의 분류 결과에 기초하여, 입력된 샘플에 대한 종을 식별할 수 있다.
- [150] 본 개시의 예시들은 임상적으로 중요한 마이크로박테리아를 정확하게 분류하는 방안에 대해서 주로 설명하였다. 그러나, 본 개시의 범위가 이에 제한되는 것은 아니며, 본 개시에 따른 적어도 네거티브 마커를 이용하는 머신러닝 기법은, 샘플을 유사한 그룹 중에서 분류하는 다양한 목적을 위해서 사용될 수 있다. 즉, 본 개시에 따른 포지티브 마커 및 네거티브 마커를 추출하는 특징과, 포지티브

마커 및 네거티브 마커에 기초한 머신 러닝 분류자에 대한 특징은, 샘플을 유사 그룹 중에서 정확하게 분류하는 다양한 기술에 적용될 수 있다.

- [151] 본 개시에 따라서 포지티브 마커 및 네거티브 마커를 TF-IDF 방식으로 추출하여 머신 러닝의 특징으로 이용하고, 특히 네거티브 마커를 유사 종 분류 및 종 식별에 대해 적용함으로써, 머신 러닝 기법에 무관하게 다양한 방식의 머신 러닝 기법의 분류 성능을 높일 수 있다. 또한, 본 개시에 따르면 유사 종 분류에 있어서의 CCI 계산과 마커-기반 머신 러닝 분류자를 조합함으로써, CCI 계산만으로는 정확하게 분류할 수 없었던 유사 종을 보다 정확하게 분류할 수 있다.
- [152] 본 개시의 예시적인 방법들은 설명의 명확성을 위해서 동작의 시리즈로 표현되어 있지만, 이는 단계가 수행되는 순서를 제한하기 위한 것은 아니며, 필요한 경우에는 각각의 단계가 동시에 또는 상이한 순서로 수행될 수도 있다. 본 개시에 따른 방법을 구현하기 위해서, 예시하는 단계에 추가적으로 다른 단계를 포함하거나, 일부의 단계를 제외하고 나머지 단계를 포함하거나, 또는 일부의 단계를 제외하고 추가적인 다른 단계를 포함할 수도 있다.
- [153] 본 개시의 다양한 실시예는 모든 가능한 조합을 나열한 것이 아니고 본 개시의 대표적인 양상을 설명하기 위한 것이며, 다양한 실시예에서 설명하는 사항들은 독립적으로 적용되거나 또는 둘 이상의 조합으로 적용될 수도 있다.
- [154] 또한, 본 개시의 다양한 실시예는 하드웨어, 펌웨어(firmware), 소프트웨어, 또는 그들의 결합 등에 의해 구현될 수 있다. 하드웨어에 의한 구현의 경우, 하나 또는 그 이상의 ASICs(Application Specific Integrated Circuits), DSPs(Digital Signal Processors), DSPDs(Digital Signal Processing Devices), PLDs(Programmable Logic Devices), FPGAs(Field Programmable Gate Arrays), 범용 프로세서(general processor), 컨트롤러, 마이크로 컨트롤러, 마이크로 프로세서 등에 의해 구현될 수 있다.
- [155] 본 개시의 범위는 다양한 실시예의 방법에 따른 동작이 장치 또는 컴퓨터 상에서 실행되도록 하는 소프트웨어 또는 머신-실행가능한 명령들(예를 들어, 운영체제, 애플리케이션, 펌웨어(firmware), 프로그램 등), 및 이러한 소프트웨어 또는 명령 등이 저장되어 장치 또는 컴퓨터 상에서 실행 가능한 비-일시적 컴퓨터-판독가능 매체(non-transitory computer-readable medium)를 포함한다.

산업상 이용가능성

- [156] 본 개시의 실시예들은 머신 러닝에 기반한 다양한 분석 방법 및 장치에 적용될 수 있다.

청구범위

- [청구항 1] 유사 종을 식별하는 방법에 있어서,
 입력된 샘플에 대한 제 1 질량 정보를 추출하는 단계;
 상기 제 1 질량 정보에 기초하여, 적어도 네거티브 마커에 기반한 머신 러닝 모델을 이용하여 상기 입력된 샘플을 분류하는 단계; 및
 상기 분류 결과에 기초하여 상기 입력된 샘플에 대한 종을 식별하는 단계를 포함하는,
 유사 종 식별 방법.
- [청구항 2] 제 1 항에 있어서,
 상기 분류하는 단계는,
 포지티브 마커 및 상기 네거티브 마커를 이용하여 상기 입력된 샘플을 분류하는 것을 포함하는,
 유사 종 식별 방법.
- [청구항 3] 제 2 항에 있어서,
 상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 유사 종에 속한 샘플들의 각각에 대해서 미리 추출되는,
 유사 종 식별 방법.
- [청구항 4] 제 2 항에 있어서,
 상기 포지티브 마커는 대립 종에 비하여 타겟 종에서 빈번하게 나타나는 질량 정보를 포함하는,
 유사 종 식별 방법.
- [청구항 5] 제 2 항에 있어서,
 상기 네거티브 마커는 타겟 종에 비하여 대립 종에서 빈번하게 나타나는 질량 정보를 포함하는,
 유사 종 식별 방법.
- [청구항 6] 제 2 항에 있어서,
 상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 유사 종에 속한 샘플들의 각각에 대한 질량 스펙트럼에 대해서 설정된 빈에 기초하여 추출되는,
 유사 종 식별 방법.
- [청구항 7] 제 6 항에 있어서,
 상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 질량 스펙트럼의 피크 값이 위치하는 빈의 번호의 집합으로 표현되는,
 유사 종 식별 방법.
- [청구항 8] 제 6 항에 있어서,
 하나의 빈은 하나 이상의 다른 빈과 일부 중첩되는,
 유사 종 식별 방법.

[청구항 9] 제 6 항에 있어서,
상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 질량 스펙트럼의 피크 값이 위치하는 bin의 bin도 정보에 기초하여 계산되는, 유사 종 식별 방법.

[청구항 10] 제 9 항에 있어서,
상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 bin bin도 정보에 대해서 TF-IDF(Term Frequency-Inverse Document Frequency) 계산에 기초하여 추출되는, 유사 종 식별 방법.

[청구항 11] 제 10 항에 있어서,
상기 포지티브 마커는, 수학적

$$TF - IDF_{bin(i)} = \frac{F_{bin(i),sample_t}}{N_t} \times \log \left(\frac{N_o}{F_{bin(i),sample_o}} \right)$$

에 기초하여 계산되며,
상기 수학적식에서 t는 타겟 종을 의미하고, o는 대립 종을 의미하고, Nt는 타겟 종에 대한 전체 개수를 의미하고, No는 대립 종에 대한 전체 개수를 의미하며, Fbin(i)는 i-번째 bin에 대한 카운트 값을 의미하는, 유사 종 식별 방법.

[청구항 12] 제 11 항에 있어서,
상기 포지티브 마커는, 상기 수학적식에 의해 계산된 TF-IDF 값이 소정의 임계치를 초과하는 경우에 상기 포지티브 마커로서 설정되는, 유사 종 식별 방법.

[청구항 13] 제 10 항에 있어서,
상기 네거티브 마커는, 수학적

$$TF - IDF_{bin(i)} = \frac{F_{bin(i),sample_o}}{N_o} \times \log \left(\frac{N_t}{F_{bin(i),sample_t}} \right)$$

에 기초하여 계산되며,
상기 수학적식에서 t는 타겟 종을 의미하고, o는 대립 종을 의미하고, Nt는 타겟 종에 대한 전체 개수를 의미하고, No는 대립 종에 대한 전체 개수를 의미하며, Fbin(i)는 i-번째 bin에 대한 카운트 값을 의미하는, 유사 종 식별 방법.

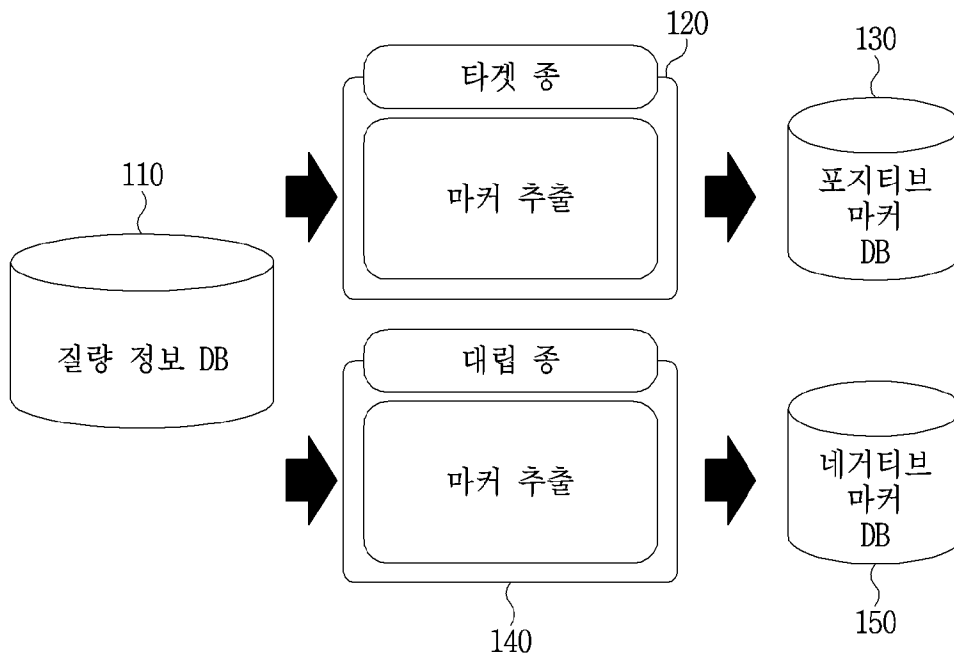
[청구항 14] 제 13 항에 있어서,
상기 네거티브 마커는, 상기 수학적식에 의해 계산된 TF-IDF 값이 소정의 임계치를 초과하는 경우에 상기 네거티브 마커로서 설정되는, 유사 종 식별 방법.

[청구항 15] 제 2 항에 있어서,

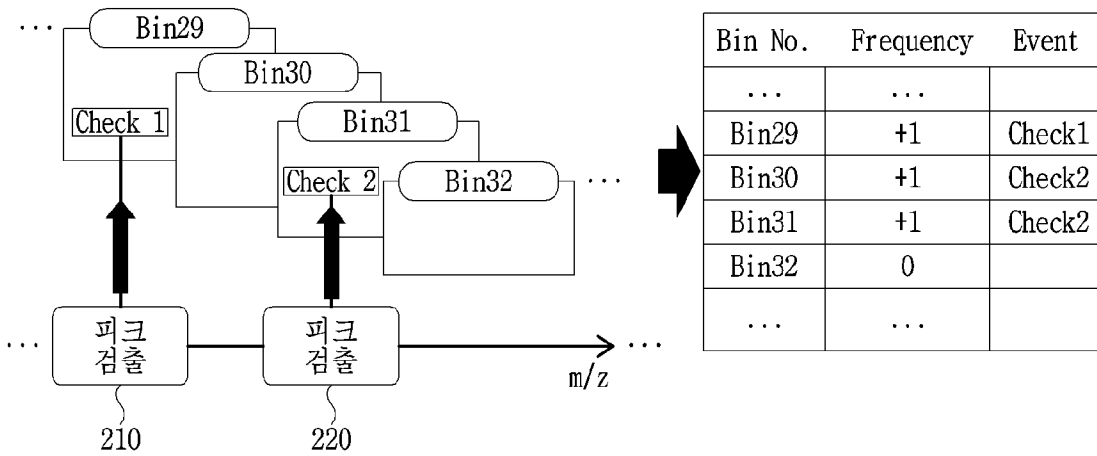
상기 포지티브 마커 및 상기 네거티브 마커의 각각은, 상기 머신 러닝 모델의 학습을 위한 특징 추출을 위한 전처리 단계로서 생성되는, 유사 종 식별 방법.

- [청구항 16] 제 1 항에 있어서,
상기 분류하는 단계는,
상기 제 1 질량 정보와, 하나 이상의 샘플의 각각에 대해서 미리 저장된 제 2 질량 정보에 기초하여 CCI(Composite Correlation Index)를 계산하는 단계; 및
계산된 상기 CCI에 기초하여 상기 분류에 대한 후보를 결정하는 단계를 더 포함하는,
유사 종 식별 방법.
- [청구항 17] 유사 종을 식별하는 장치에 있어서,
입력된 샘플에 대한 제 1 질량 정보를 추출하는 질량 분석기; 및
상기 제 1 질량 정보에 기초하여, 적어도 네거티브 마커 데이터베이스에 저장된 네거티브 마커에 기반한 머신 러닝 모델을 이용하여 상기 입력된 샘플을 분류하는 분류기를 포함하고,
상기 장치는 상기 분류 결과에 기초하여 상기 입력된 샘플에 대한 종을 식별하는,
유사 종 식별 장치.
- [청구항 18] 제 17 항에 있어서,
상기 분류기는,
포지티브 마커 데이터베이스에 저장된 포지티브 마커 및 상기 네거티브 마커를 이용하여 상기 입력된 샘플을 분류하는,
유사 종 식별 장치.
- [청구항 19] 제 18 항에 있어서,
상기 포지티브 마커 데이터베이스 및 상기 네거티브 마커 데이터베이스는, 상기 유사 종에 속한 샘플들의 각각에 대한 질량 스펙트럼에 대해서 설정된 빈에 기초하여 추출된 포지티브 마커 및 네거티브 마커를 각각 저장하는,
유사 종 식별 장치.
- [청구항 20] 제 17 항에 있어서,
상기 장치는,
상기 제 1 질량 정보와, 하나 이상의 샘플의 각각에 대해서 데이터베이스에 미리 저장된 제 2 질량 정보에 기초하여 CCI(Composite Correlation Index)를 계산하고, 계산된 상기 CCI에 기초하여 상기 분류에 대한 후보를 결정하는 유사도 계산기를 더 포함하는,
유사 종 식별 장치.

[도1]



[도2]

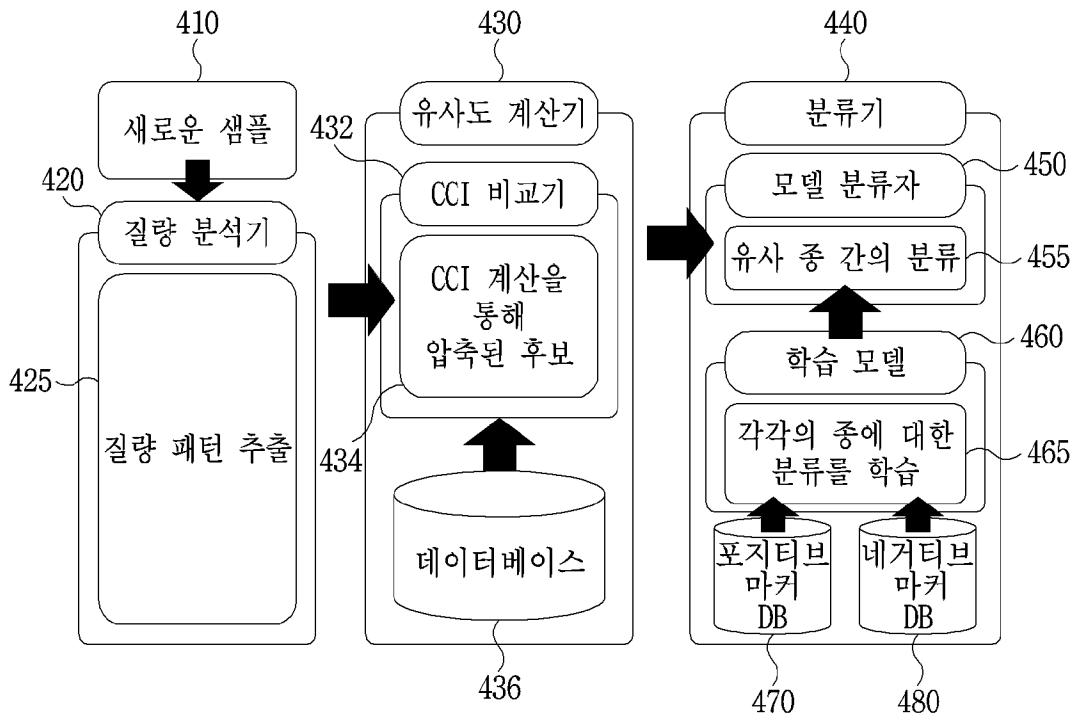


[도3]

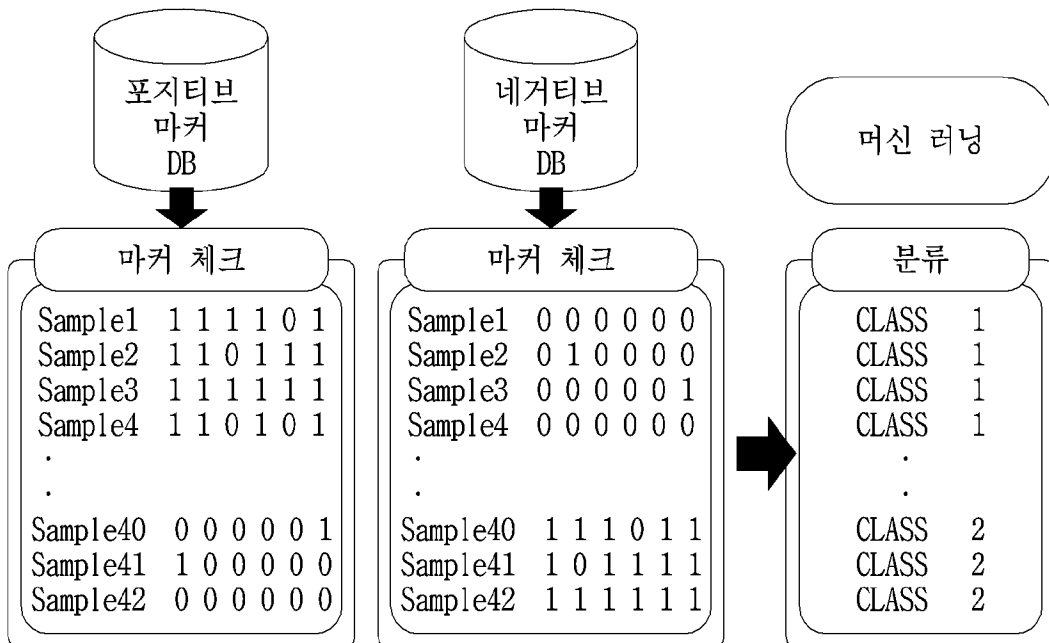
Positive Marker DB	
Bacteria_ID	Peak_Set
a1	bin1,bin31,bin42,...
a2	bin3,bin25,bin49,...
a3	bin9,bin44,bin47,...
a4	bin12,bin22,bin33,...
...	...

Negative Marker DB	
Bacteria_ID	Peak_Set
a1	bin7,bin35,bin49,...
a2	bin6,bin33,bin88,...
a3	bin17,bin36,bin41,...
a4	bin19,bin39,bin47,...
...	...

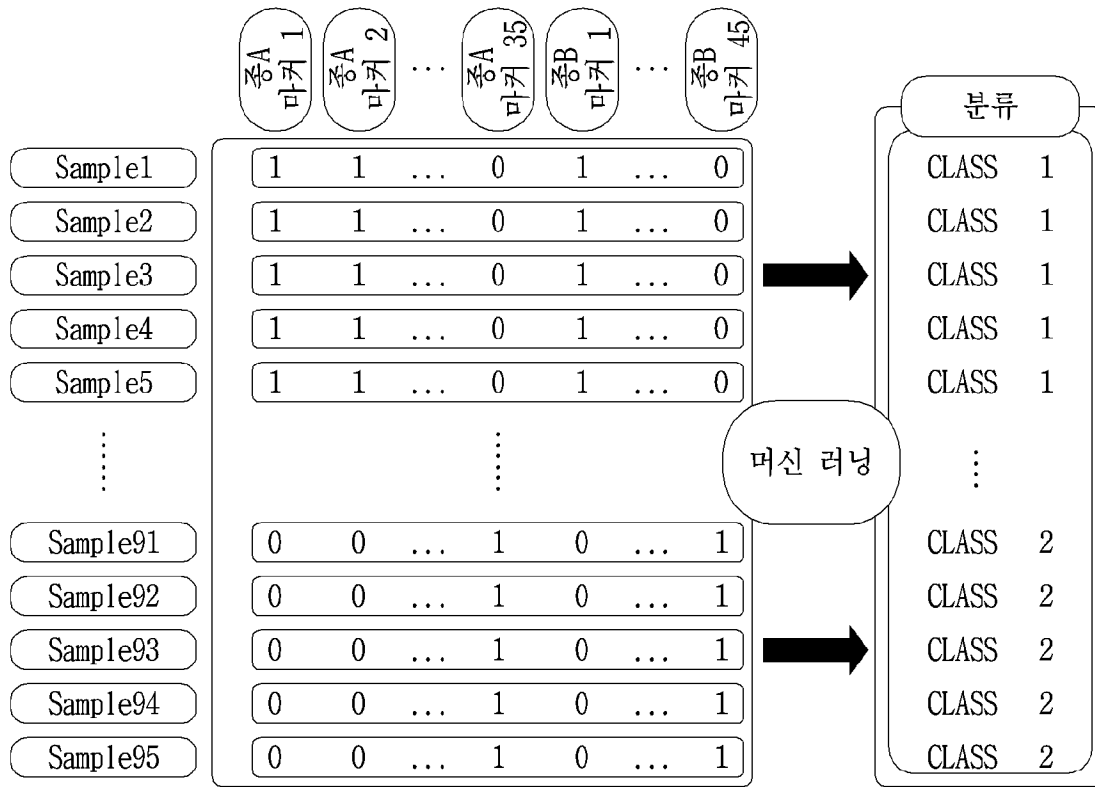
[도4]



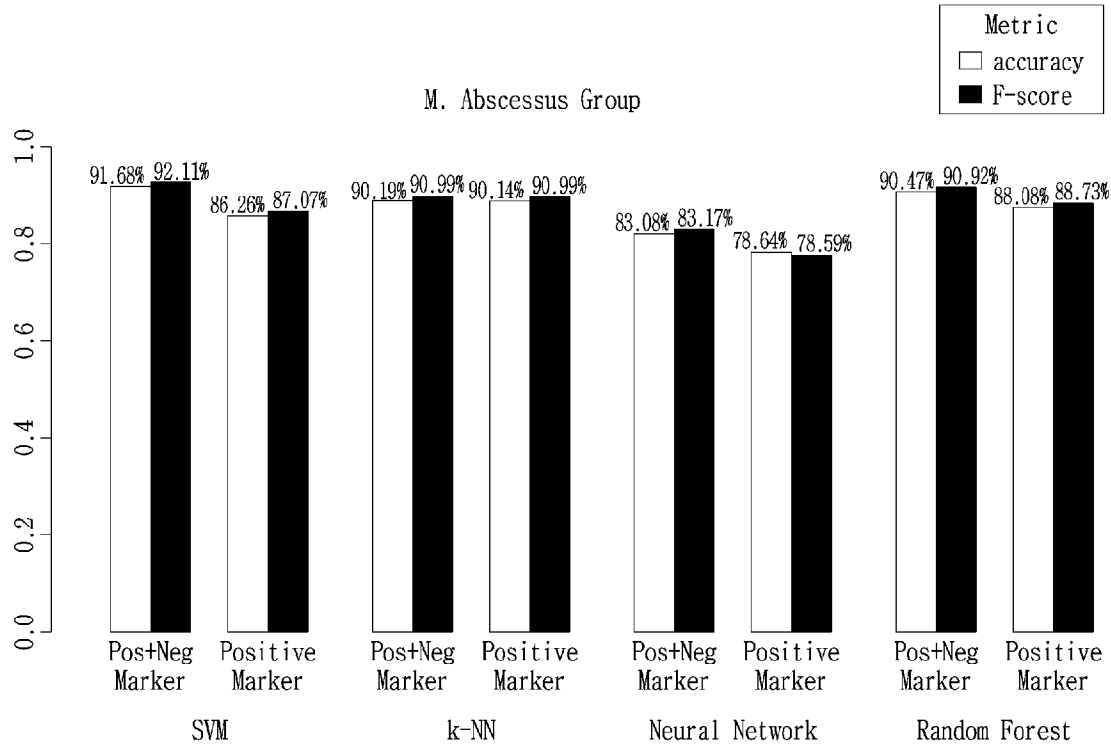
[도5]



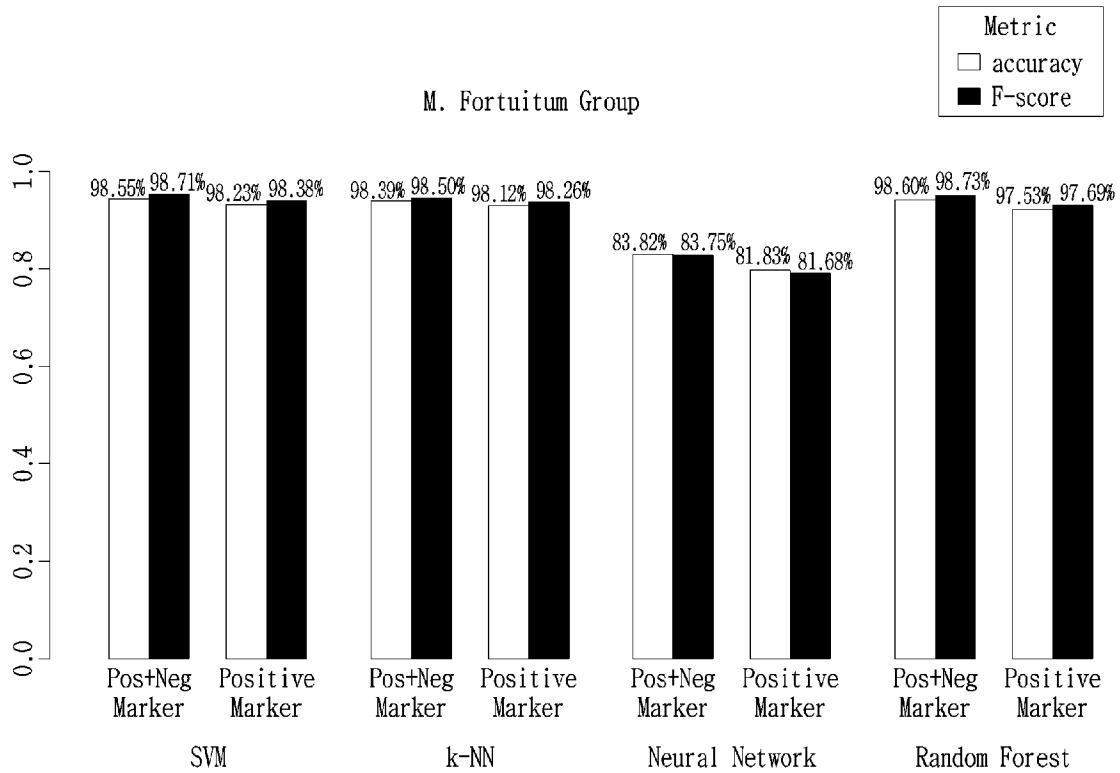
[도6]



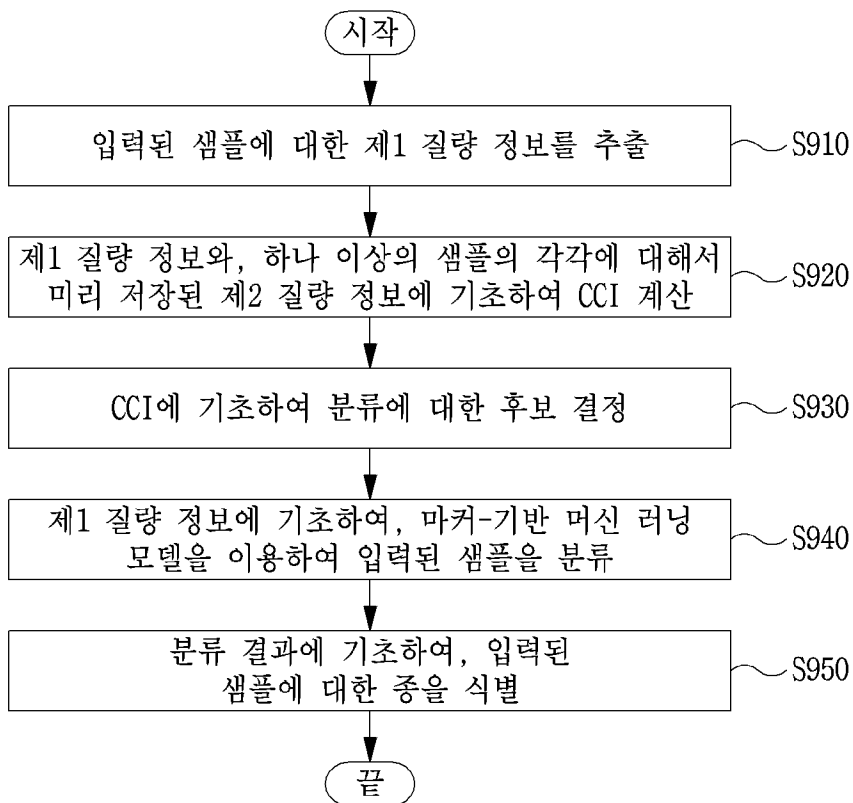
[도7]



[도8]



[도9]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2018/006892

A. CLASSIFICATION OF SUBJECT MATTER

G01N 33/68(2006.01)i, G06F 19/22(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)

G01N 33/68; G01N 27/62; C07K 7/06; H01J 49/26; H01J 49/00; G06F 19/22; 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 aboveElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS (KIPO internal) & Keywords: positive marker, negative marker, identification, classification, sample, mass

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2012-0197535 A1 (GOODLETT et al.) 02 August 2012 See paragraphs [0004]-[0021], [0048]-[0093], [0134]-[0160], [0235]-[0238]; and figures 1, 3A, 4-6.	1-9,15,17-19
Y		16,20
A		10-14
Y	CN 105116078 A (INSTITUTE OF TROPICAL BIOSCIENCE AND BIOTECHNOLOGY et al.) 02 December 2015 See paragraph [0084]; and table 1.	16,20
A	US 2012-0264156 A1 (BEAULIEU et al.) 18 October 2012 See paragraphs [0079]-[0096], [0202]-[0220]; and tables 4A-4B.	1-20
A	JP 2014-514566 A (BIOMERIEUX, INC.) 19 June 2014 See paragraphs [0040]-[0043], [0049]-[0051], [0177]-[0179]; and tables 28, 31, 34.	1-20
A	US 2005-0230615 A1 (FURUTANI et al.) 20 October 2005 See paragraphs [0059]-[0064], [0065]-[0066]; and figures 4, 7-8.	1-20

 Further documents are listed in the continuation of Box C.
 See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	


Date of the actual completion of the international search

21 SEPTEMBER 2018 (21.09.2018)

Date of mailing of the international search report

21 SEPTEMBER 2018 (21.09.2018)

Name and mailing address of the ISA/KR


 Korean Intellectual Property Office
 Government Complex Daejeon Building 4, 189, Cheongsa-ro, Seo-gu,
 Daejeon, 35208, 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/KR2018/006892

Patent document cited in search report	Publication date	Patent family member	Publication date
US 2012-0197535 A1	02/08/2012	US 9273339 B2	01/03/2016
CN 105116078 A	02/12/2015	NONE	
US 2012-0264156 A1	18/10/2012	BR 112012008537 A2	29/11/2016
		CN 102656276 A	05/09/2012
		CN 102656276 B	06/07/2016
		EP 2488660 A2	22/08/2012
		EP 2488660 B1	10/08/2016
		FR 2951548 A1	22/04/2011
		FR 2951548 B1	11/11/2011
		JP 2013-508675 A	07/03/2013
		JP 5754742 B2	29/07/2015
		KR 10-1833419 B1	28/02/2018
		KR 10-2012-0102634 A	18/09/2012
		MX 2012004310 A	08/05/2012
		MX 341204 B	11/08/2016
		RU 2012-119798 A	20/11/2013
		RU 2604199 C2	10/12/2016
		WO 2011-045544 A2	21/04/2011
		WO 2011-045544 A3	11/08/2011
JP 2014-514566 A	19/06/2014	CA 2833456 A1	26/10/2012
		CN 103765216 A	30/04/2014
		CN 103765216 B	11/08/2017
		CN 107271686 A	20/10/2017
		EP 2702407 A2	05/03/2014
		EP 2702407 B1	06/06/2018
		JP 6216711 B2	18/10/2017
		US 2015-0031063 A1	29/01/2015
		US 2017-0082636 A1	23/03/2017
		US 9551020 B2	24/01/2017
		US 9874568 B2	23/01/2018
		WO 2012-143535 A2	26/10/2012
		WO 2012-143535 A3	13/12/2012
US 2005-0230615 A1	20/10/2005	CA 2552005 A1	21/07/2005
		EP 1700327 A2	13/09/2006
		EP 1700327 A4	05/05/2010
		US 7170052 B2	30/01/2007
		WO 2005-065307 A2	21/07/2005
		WO 2005-065307 A3	25/01/2007

A. 발명이 속하는 기술분류(국제특허분류(IPC))
G01N 33/68(2006.01)i, G06F 19/22(2011.01)i, G06N 3/08(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)
G01N 33/68; G01N 27/62; C07K 7/06; H01J 49/26; H01J 49/00; G06F 19/22; 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) & 키워드:포지티브 마커, 네거티브 마커, 식별, 분류, 샘플, 질량

C. 관련 문헌

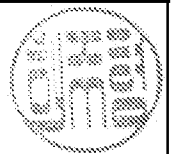
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	US 2012-0197535 A1 (GOODLETT 등) 2012.08.02 단락 [0004]-[0021], [0048]-[0093], [0134]-[0160], [0235]-[0238]; 및 도면 1, 3A, 4-6 참조.	1-9, 15, 17-19
Y		16, 20
A		10-14
Y	CN 105116078 A (INSTITUTE OF TROPICAL BIOSCIENCE AND BIOTECHNOLOGY 등) 2015.12.02 단락 [0084]; 및 표 1 참조.	16, 20
A	US 2012-0264156 A1 (BEAULIEU 등) 2012.10.18 단락 [0079]-[0096], [0202]-[0220]; 및 표 4A-4B 참조.	1-20
A	JP 2014-514566 A (BIOMERIEUX, INC.) 2014.06.19 단락 [0040]-[0043], [0049]-[0051], [0177]-[0179]; 및 표 28, 31, 34 참조.	1-20
A	US 2005-0230615 A1 (FURUTANI 등) 2005.10.20 단락 [0059]-[0064], [0065]-[0066]; 및 도면 4, 7-8 참조.	1-20

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 09월 21일 (21.09.2018)	국제조사보고서 발송일 2018년 09월 21일 (21.09.2018)
--	---

ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578	심사관 허주형 전화번호 +82-42-481-8150
---	------------------------------------



국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
US 2012-0197535 A1	2012/08/02	US 9273339 B2	2016/03/01
CN 105116078 A	2015/12/02	없음	
US 2012-0264156 A1	2012/10/18	BR 112012008537 A2 CN 102656276 A CN 102656276 B EP 2488660 A2 EP 2488660 B1 FR 2951548 A1 FR 2951548 B1 JP 2013-508675 A JP 5754742 B2 KR 10-1833419 B1 KR 10-2012-0102634 A MX 2012004310 A MX 341204 B RU 2012-119798 A RU 2604199 C2 WO 2011-045544 A2 WO 2011-045544 A3	2016/11/29 2012/09/05 2016/07/06 2012/08/22 2016/08/10 2011/04/22 2011/11/11 2013/03/07 2015/07/29 2018/02/28 2012/09/18 2012/05/08 2016/08/11 2013/11/20 2016/12/10 2011/04/21 2011/08/11
JP 2014-514566 A	2014/06/19	CA 2833456 A1 CN 103765216 A CN 103765216 B CN 107271686 A EP 2702407 A2 EP 2702407 B1 JP 6216711 B2 US 2015-0031063 A1 US 2017-0082636 A1 US 9551020 B2 US 9874568 B2 WO 2012-143535 A2 WO 2012-143535 A3	2012/10/26 2014/04/30 2017/08/11 2017/10/20 2014/03/05 2018/06/06 2017/10/18 2015/01/29 2017/03/23 2017/01/24 2018/01/23 2012/10/26 2012/12/13
US 2005-0230615 A1	2005/10/20	CA 2552005 A1 EP 1700327 A2 EP 1700327 A4 US 7170052 B2 WO 2005-065307 A2 WO 2005-065307 A3	2005/07/21 2006/09/13 2010/05/05 2007/01/30 2005/07/21 2007/01/25