(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0048991
(43) 공개일자 2024년04월16일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6883 (2018.01) *C12Q 1/689* (2018.01)
G16B 40/00 (2019.01) *G16H 50/20* (2018.01)
(52) CPC특허분류
C12Q 1/6883 (2022.01)
C12Q 1/689 (2018.05)
(21) 출원번호 10-2022-0128962
(22) 출원일자 2022년10월07일
심사청구일자 2022년10월07일

(71) 출원인
한림대학교 산학협력단
강원도 춘천시 한림대학길 1, 한림대학교(옥천동)
(72) 발명자
석기태
강원도 원주시 늘품로 199, 109동 1032호(반곡동, 원주반곡아이파크)
이은주
강원도 춘천시 국사봉길 7, 104동 1403호 (퇴계동)
(뒷면에 계속)
(74) 대리인
특허법인본

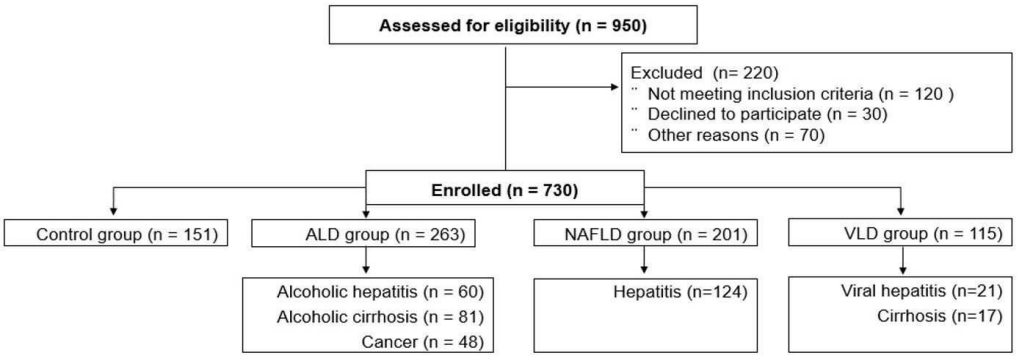
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **장내 미생물 균총 분석을 통한 비알코올성 간질환 감별 진단 방법**

(57) 요 약

본 발명은 시료의 16S Ribosomal RNA sequencing 결과 정보를 바탕으로, 머신 러닝 기법 중 벡터 머신 (Vector machine), 랜덤포레스트 (Random forest), 다층 퍼셉트론 (Multi-layer perceptron), 합성곱 신경망 (Convolution neural networks)를 적용하여 비알코올성 간질환을 예측하는 바이오마커를 발굴하고, 이를 비알코올성 간질환의 진단을 위한 정보제공방법에 제공하는 것이다. 본 발명은 (1) 개체의 생물학적 시료의 16S rRNA 유전 정보를 이용하여 미생물 균총 분석을 수행하는 단계; (2) 상기 미생물 균총 분석 결과를 검사 대상 객체로부터 수득한 생물학적 시료의 미생물 균총과 비교하여, 비알코올성 간염으로 진단하는, 알코올성 간질환 진단을 위한 정보 제공 방법을 제공한다.

대 표 도 - 도1a



(52) CPC특허분류
G16B 40/00 (2019.02)
G16H 50/20 (2018.01)
(72) 발명자
박인규
강원도 춘천시 향교뒷길 15 (교동)
오기광
강원도 춘천시 퇴계로 168, 215동 1403호
원성민
강원도 춘천시 후석로 325 춘천더샵아파트 110동 1405호
윤상준
강원도 춘천시 전원안길33번길 11, 401호 (교동)

이 발명을 지원한 국가연구개발사업
과제고유번호 1345332092
과제번호 2020R1A6A1A03043026
부처명 교육부
과제관리(전문)기관명 한국연구재단
연구사업명 중점연구소지원사업
연구과제명 장내 마이크로바이옴을 이용한 간질환 진단 및 치료기술 개발
기 여 율 1/2
과제수행기관명 한림대학교
연구기간 2020.06.01 ~ 2029.05.31
이 발명을 지원한 국가연구개발사업
과제고유번호 0020622
과제번호 G02F09271784901
부처명 중소벤처기업부
과제관리(전문)기관명 한국산업기술진흥원
연구사업명 규제자유특구혁신사업육성(R&D)
연구과제명 만성(알콜성) 간질환 진단·예측 정밀의료 AI솔루션 개발 및 실증
기 여 율 1/2
과제수행기관명 (주)아이도트
연구기간 2022.01.01 ~ 2023.11.30

명 세 서

청구범위

청구항 1

(1) 개체의 생물학적 시료의 16S rRNA 유전 정보를 이용하여 미생물 균총 분석을 수행하는 단계; 및

(2) 상기 미생물 균총 분석 결과를 검사 대상 객체로부터 수득한 생물학적 시료의 미생물 균총과 비교하여, 비알코올성 간염으로 진단하는 단계;를 포함하는, 비알코올성 간질환 진단을 위한 정보 제공 방법.

청구항 2

제 1 항에 있어서,

상기 (2) 단계에서 미생물 균총 분석 결과의 비알코올성 간질환 대조군과 비교하여,

검사 대상 객체로부터 수득한 액티노박테리아(*Actinobacteria*), 피르미쿠테스(*Firmicutes*) 및 TM7으로 이루어진 군에서 선택된 1종 이상의 문(phylum) 미생물의 분포가 높은 경우,

비알코올성 간염으로 진단하는 것을 특징으로 하는, 비알코올성 간질환 진단을 위한 정보 제공 방법.

청구항 3

제 1 항에 있어서,

상기 (2) 단계에서 미생물 균총 분석 결과의 비알코올성 간질환 대조군과 비교하여,

검사 대상 객체로부터 수득한 액티노박테리아(*Actinobacteria*), 코리오박테리아(*Coriobacteriia*), 바실리(*Bacilli*), 에리시페로트리치(*Erysipelotrichi*) 및 TM7_3으로 이루어진 군에서 선택된 1종 이상의 강(class) 미생물의 분포가 높은 경우,

비알코올성 간염으로 진단하는 것을 특징으로 하는, 비알코올성 간질환 진단을 위한 정보 제공 방법.

청구항 4

제 1 항에 있어서,

상기 (2) 단계에서 미생물 균총 분석 결과의 비알코올성 간질환 대조군과 비교하여,

검사 대상 객체로부터 수득한 비피도박테리아레스(*Bifidobacteriales*), 코리오박테리아레스(*Coriobacteriales*) 및 에리시페로트리차레스(*Erysipelotrichales*)로 이루어진 군에서 선택된 1종 이상의 목(order) 미생물의 분포가 높은 경우,

비알코올성 간염으로 진단하는 것을 특징으로 하는, 비알코올성 간질환 진단을 위한 정보 제공 방법.

청구항 5

제 1 항에 있어서,

상기 (2) 단계에서 미생물 균총 분석 결과의 비알코올성 간질환 대조군과 비교하여,

검사 대상 객체로부터 수득한 비피도박테리아시에(*Bifidobacteriaceae*), 아이로콕카시에(*Aerococcaceae*), 락토바실라시에(*Lactobacillaceae*) 및 스트렙토코카시에(*Streptococcaceae*)로 이루어진 군에서 선택된 1종 이상의 과(family) 미생물의 분포가 높은 경우,

비알코올성 간염으로 진단하는 것을 특징으로 하는, 비알코올성 간질환 진단을 위한 정보 제공 방법.

청구항 6

제 1 항에 있어서,

상기 (2) 단계에서 미생물 균총 분석 결과의 비알코올성 간질환 대조군과 비교하여,

검사 대상 객체로부터 수득한 콜린셀라(*Collinsella*), 그라눌리카텔라(*Granulicatella*), 웨이셀라(*Weissella*), 스트렙토코쿠스(*Streptococcus*), 부티리시코쿠스(*Butyricicoccus*), 서브돌리그라눌럼(*Subdoligranulum*), 아네로시누스(*Anaerosinus*), 숙시니스피라(*Succinispira*), 카테니박테리움(*Catenibacterium*) 및 홀데마니아(*Holdemania*)로 이루어진 군에서 선택된 1종 이상의 속(genus) 미생물의 분포가 높은 경우,

비알코올성 간염으로 진단하는 것을 특징으로 하는, 비알코올성 간질환 진단을 위한 정보 제공 방법.

청구항 7

제 1 항에 있어서,

상기 (2) 단계에서 미생물 균총 분석 결과의 비알코올성 간질환 대조군과 비교하여,

검사 대상 객체로부터 수득한 코프로피루스(*Coprophilus*), 인디스틴크투스(*Indistinctus*), 발라에노프테라에(*Balaenopterae*), 헬레니카(*Hellenica*), 이콰이(*Equi*), 볼티에(*Bolteae*), 풀리카에코룸(*Pullicaecorum*), 바리아빌레(*Variabile*), 글리세리니(*Glycerini*), 모빌리스(*Mobilis*), 비포르메(*Biforme*) 및 미트수오카이(*Mitsuokai*)로 이루어진 군에서 선택된 1종 이상의 종(species) 미생물의 분포가 높은 경우,

비알코올성 간염으로 진단하는 것을 특징으로 하는, 비알코올성 간질환 진단을 위한 정보 제공 방법.

청구항 8

전자 장치의 프로세서에 의해 실행되어, 상기 전자 장치로 하여금,

개체의 생물학적 시료의 16S rRNA 유전 정보를 이용하여 미생물 균총 분석을 수행하는 단계; 및

상기 미생물 균총 분석 결과를 검사 대상 객체로부터 수득한 생물학적 시료의 미생물 균총과 비교하여, 비알코올성 간염 여부를 진단하는 단계;를 포함하는 제어 방법을 수행하도록 하는, 컴퓨터 판독 가능 매체에 저장된 컴퓨터 프로그램.

청구항 9

전자 장치에 있어서,

비알코올성 간질환의 분류를 위한 인공지능 모델을 포함하는 메모리; 및

개체의 생물학적 시료의 16S rRNA 유전 정보에 대한 미생물 균총 분석의 결과 정보를 상기 인공지능 모델에 입력하여 비알코올성 간염 여부에 대한 진단을 수행하는, 프로세서;를 포함하고,

상기 인공지능 모델은,

상기 결과 정보 내에서 축소된 특징 정보를 추출하는 특징 추출 모듈; 및

상기 특징 정보를 바탕으로 비알코올성 간염 여부에 대한 진단을 수행하는 간질환 분류 모듈;을 포함하는, 전자 장치.

청구항 10

비알코올성 간질환의 분류를 위한 인공지능 모델을 포함하는 전자 장치의 제어 방법에 있어서,

상기 전자 장치가 개체의 생물학적 시료의 16S rRNA 유전 정보에 대한 미생물 균총 분석의 결과 정보를 상기 인공지능 모델에 입력하는 단계; 및상기 전자 장치가 상기 인공지능 모델의 출력에 따라 비알코올성 간염 여부에 대한 진단을 수행하는 단계;를 포함하는, 전자 장치의 제어 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 장내 미생물 균총 분석을 통한 비알코올성 간질환 감별 진단 방법에 관한 것으로, 보다 구체적으로는

머신 러닝 기법을 기반으로 장내 미생물 균총 분석을 하여 질병 분류군을 정확하게 설명하는 미생물 균총의 다양성(종, 또는 속 이상의 분류) 정보를 확립하고, 이를 이용하여 비알코올성 간질환 여부를 진단을 위한 정보 제공 방법에 관한 것이다.

배 경 기 술

[0002] 간은 생체에서 단백질 합성, 포도당 및 기타 탄수화물 대사, 핵산 대사, 지질 대사, 약물 대사 등을 담당하는 중요 장기로, 생명 유지에 필수적인 장기 중 하나이다. 따라서 간질환은 일반인에게 치명적으로 작용할 수 있다.

[0003] 만성 간질환 (Chronic Liver Disease)은 전 세계적으로 높은 유병률과 사망률을 보이는 질환으로서, 알코올성 간질환 (Alcoholic Liver Disease), 비알코올성 간질환 (Non-Alcoholic Liver Disease), 바이러스성 간질환 (Viral Liver Disease)으로 분류된다. 병리학적으로 만성 간질환은 지방증, 염증, 섬유증, 간암 및 간 재생까지 포함된다. 세포 외 기질 단백질의 생성과 분해의 항상성이 지속적으로 무너지면 섬유성 상처 (Scars)가 발생되어, 간 조직에 구조적으로 이상(異常)이 생기며, 간경화 (Liver cirrhosis) 및 간암 (Liver cancer)로 이어진다. 간경화는 만성 간 손상으로 발생하며, 재생성 결절에 섬유 띠로 둘러쌓여 간섬유화로 발전된다. 그러나, 그 진행 발병인자가 없어지면 가역적으로 회복될 수 있다.

[0004] 인간의 장내 미생물군 유전체는 박테리아, 균류, 및 바이러스를 포함한 다양한 미생물로 구성된 하나의 생태계이다. 장내 미생물은 인간 건강의 면역 체계, 염증 반응, 영양 균형, 항상성 유지, 그리고 순환 및 신진 대사에서 여러 역할을 한다. 또한, 장내 미생물은 다양한 질병의 바이오 마커로 주목받고 있다.

[0005] 인간과 장내 미생물의 균형이 무너지게 되면, 미생물의 구성과 기능이 변하게 되는데 이를 장내세균불균형 (Dysbiosis) 라고 한다. 장내미생물의 불균형은 다양한 질병의 발병과 연관이 있으며, 각 질병은 장내미생물과 깊은 상관 관계가 있음이 확인되고 있다. 만성 간질환에서 잠재적인 진단 양식, 요법 또는 예후 마커로서 장내 미생물 유전체의 역할에 대한 관심이 증가하고 있다. 현재까지, 장내 미생물의 분포는 유전적 요인, 영양, 환경, 운동, 병리, 생활 습관 등 여러 요인에 의해 영향을 받는다. 즉, 개개인의 차이와 다양성이 존재하기 때문에, 개별적인 특성과 규칙을 파악하기 어렵다. 최근 연구에 따르면 장내 미생물 균총의 구성이 만성 간질환 진행 경과에 따라 다르다는 것이 입증됨으로써, 장내미생물 유전체 분포를 이용하여 간질환을 진단하는 것이 가능하다.

[0006] 환자 중심의 정확하고 맞춤형 의료를 구현하기 위해서는 인공지능 (Artificial Intelligence)를 활용한 진단 기술의 적용이 요구되어진다. 최근 인공지능 기반 간질환 진단과 관련된 데이터가 공개되고 있지만, 간질환의 다양한 질병군에 대한 정제된 데이터는 부족한 실정이다. 그러나, 차선책으로 인공지능의 한 분야인 머신 러닝 (Machine learning)은 악성 종양 (Malignancy), 대사 질환 (Metabolic disorder), 그리고 장 질환 (Intestinal disease) 등 다양한 질병 진단에 활용되고 있다. 이에, 본 발명자들은 분변의 16S Ribosomal RNA sequencing 정보를 고도화된 머신 러닝 기법으로 분석하여, 정확하고 효율적인 비알코올성 간질환 진단법을 구축하고자 하였다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 한국 등록특허공보 제10-2373886호

(특허문헌 0002) 한국 공개특허공보 제10-2022-0068865호

발명의 내용

해결하려는 과제

[0008] 본 발명은 시료의 16S Ribosomal RNA sequencing 결과 정보를 바탕으로, 머신 러닝 기법 중 벡터 머신 (Vector machine), 랜덤포레스트 (Random forest), 다층 퍼셉트론 (Multi-layer perceptron), 합성곱 신경망 (Convolution neural networks)를 적용하여 비알코올성 간질환을 예측하는 바이오마커를 발굴하고, 이를 비알코올성 간질환의 진단을 위한 정보제공방법에 제공하는 것이다.

[0009] 본 개시의 목적들은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 본 개시의 다른 목적 및 장점들은 하기의 설명에 의해서 이해될 수 있고, 본 개시의 실시 예에 의해 보다 분명하게 이해될 것이다. 또한, 본 개시의 목적 및 장점들은 특허 청구 범위에 나타낸 수단 및 그 조합에 의해 실현될 수 있음을 쉽게 알 수 있을 것이다.

과제의 해결 수단

[0010] 본 발명은, (1) 개체의 생물학적 시료의 16S rRNA 유전 정보를 이용하여 미생물 균총 분석을 수행하는 단계; 및 (2) 상기 미생물 균총 분석 결과를 검사 대상 객체로부터 수득한 생물학적 시료의 미생물 균총과 비교하여, 비알코올성 간염으로 진단하는, 비알코올성 간질환 진단을 위한 정보 제공 방법을 제공한다.

[0011] 상기 (2) 단계에서 미생물 균총 분석 결과의 비알코올성 간질환 대조군과 비교하여, 검사 대상 객체로부터 수득한 액티노박테리아(*Actinobacteria*), 피르미쿠테스(*Firmicutes*) 및 TM7으로 이루어진 군에서 선택된 1종 이상의 문(phylum) 미생물; 검사 대상 객체로부터 수득한 액티노박테리아(*Actinobacteria*), 코리오박테리아(*Coriobacteriia*), 바실리(*Bacilli*), 에리시페로트리치(*Erysipelotrichi*) 및 TM7_3으로 이루어진 군에서 선택된 1종 이상의 강(class) 미생물; 검사 대상 객체로부터 수득한 비피도박테리아레스(*Bifidobacteriales*), 코리오박테리아레스(*Coriobacteriales*) 및 에리시페로트리차레스(*Erysipelotrichales*)로 이루어진 군에서 선택된 1종 이상의 목(order) 미생물; 검사 대상 객체로부터 수득한 비피도박테리아시에(*Bifidobacteriaceae*), 아이로콕카시에(*Aerococcaceae*), 락토바실라시에(*Lactobacillaceae*) 및 스트렙토코카시에(*Streptococcaceae*)로 이루어진 군에서 선택된 1종 이상의 과(family) 미생물; 검사 대상 객체로부터 수득한 콜린셀라(Collinsella), 그라눌리카텔라(Granulicatella), 웨이셀라(Weissella), 스트렙토코쿠스(Streptococcus), 부티리시코쿠스(Butyricicoccus), 서브돌리그라눌럼(Subdoligranulum), 아네로시누스(Anaerosinus), 숙시니스피라(Succinispira), 카테니박테리움(Catenibacterium) 및 홀데마니아(Holdemania)로 이루어진 군에서 선택된 1종 이상의 속(genus) 미생물; 또는 검사 대상 객체로부터 수득한 코프로피루스(Coprophilus), 인디스틴크투스(Indistinctus), 발라에노프테라에(Balaenopterae), 헬레니카(Hellenica), 이콰이(Equi), 볼티에(Bolteae), 풀리카에코룸(Pullicaecorum), 바리아빌레(Variabile), 글리세리니(Glycerini), 모빌리스(Mobilis), 비포르메(Biforme) 및 미트수오카이(Mitsuokai)로 이루어진 군에서 선택된 1종 이상의 종(species) 미생물의 분포가 높은 경우, 비알코올성 간염으로 진단할 수 있다.

[0012] 본 발명은 또한, 데이터를 입력하는 입력 수단, 입력된 데이터를 처리하는 처리 수단 및 출력 수단을 갖는 스마트폰, 태블릿, 노트북, 또는 컴퓨터에, (1) 개체의 생물학적 시료의 16S rRNA 유전 정보를 이용하여 미생물 균총 분석을 수행하는 제1 기능; 및 (2) 상기 미생물 균총 분석 결과를 검사 대상 객체로부터 수득한 생물학적 시료의 미생물 균총과 비교하여, 비알코올성 간염으로 판단하는 제2 기능을 실행시키는, 비알코올성 간질환 진단을 위한 정보 제공 프로그램을 제공할 수 있다.

발명의 효과

[0013] 본 발명에 따르면, 장내 미생물 균총 분석 및 머신 러닝 기법을 이용하여, 비알코올성 간염의 진단이 가능한 비알코올성 간질환 진단을 위한 정보 제공 방법을 제공할 수 있다.

[0014] 본 발명은 또한, 분변의 장내 미생물 균총 분석을 활용하여 머신 러닝 기법을 도입하여, 만성 간질환으로부터 간염, 간경화, 그리고 간암의 진단이 가능하여 이를 기반으로 한 체액을 활용한 진단법 개발에 기여할 수 있다.

[0015] 본 발명은 또한, 간염, 간경화, 그리고 간암의 감별 진단이 가능한 바 각 질환의 병리기전의 확인, 머신 러닝 기반 장내 미생물 균총 분석 기반 상기 질환들의 치료 및 사전 진단(Prognostics)을 위한 후속 연구 및 치료법 개발에 활용될 수 있다.

도면의 간단한 설명

[0016] 도 1a는, 환자로부터 획득한 시료 샘플을 세부 간질환으로 구분하는 과정을 나타낸 모식도이다. 도 1b는, 머신 러닝 기반으로 간질환 그룹 간 감별하는 방법을 나타낸 모식도이다.

도 2a는 알코올성 간질환 그룹에서 미생물 균총 알파 다양성을 표현한 그래프이다. 도 2b는 알코올성 간질환 그룹에서 미생물 균총 베타 다양성을 표현한 그래프이다. 도 2c는 알코올성 간질환 그룹에서 알코올성 간질환 대조군, 간염, 간경화, 그리고 간암 그룹 간 Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria, 그리고 Verrucomicrobia 의 분포도를 6가지 색상별로 시각화한 그래프이다. 도 2d는 LDA

(linear discriminant analysis) score 2 이상을 가진 분류학적 분기도 (Cladogram)를 나타낸 도이다. 도 2e는 알코올성 간질환 그룹에서 세부 그룹별 LDA effect size를 나타낸 도이다.

도 3a는 비알코올성 간질환 그룹에서 미생물 균총 알파 다양성을 표현한 그래프이다. 도 3b는 비알코올성 간질환 그룹에서 미생물 균총 베타 다양성을 표현한 그래프이다. 도 3c는 비알코올성 간질환 그룹에서 비알코올성 간질환 대조군 그리고 간염 그룹 간 *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Fusobacteria*, *Proteobacteria*, 그리고 *Verrucomicrobia* 의 분포도를 6가지 색상별로 시각화한 그래프이다. 도 3d는 LDA (linear discriminant analysis) score 2 이상을 가진 분류학적 분기도(Cladogram)를 나타낸 도이다. 도 3e는 비알코올성 간질환 그룹에서 세부 그룹별 LDA effect size를 나타낸 도이다. 도 3f는 비알코올성 간질환 그룹에서 세부 그룹별 속 (Genus) 과 종 (Species) 히트 맵 매트릭스 (Matrix)를 표현한 다이어그램이다.

도 4a는 바이러스성 간질환 그룹에서 미생물 균총 알파 다양성을 표현한 그래프이다. 도 4b는 바이러스성 간질환 그룹에서 미생물 균총 베타 다양성을 표현한 그래프이다. 도 4c는 바이러스성 간질환 그룹에서 바이러스성 간질환 대조군, 간염, 그리고 간경화 그룹 간 Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria, 그리고 Verrucomicrobia 의 분포도를 6가지 색상별로 시각화한 그래프이다. 도 4d는 LDA (linear discriminant analysis) score 2 이상을 가진 분류학적 분기도(Cladogram)를 나타낸 도이다. 도 4e는 바이러스성 간질환 그룹에서 세부 그룹별 LDA effect size를 나타낸 도이다. 도 4f는 바이러스성 간질환 그룹에서 세부 그룹별 속 (Genus) 과 종 (Species) 히트 맵 매트릭스 (Matrix)를 표현한 다이어그램이다.

도 5는 본 개시의 일 실시 예에 따른 전자 장치의 구성을 설명하기 위한 블록도,

도 6a는 알코올성 간질환 그룹 내 세부 간질환 분류를 위해 개발된 머신 러닝 기반 분류 모델 정확도 평가 결과의 ROC (Receiver operating characteristic) 그래프이다. 도 6b는 비알코올성 간질환 그룹 내 세부 간질환 분류를 위해 개발된 머신 러닝 기반 분류 모델 정확도 평가 결과의 ROC (Receiver operating characteristic) 그래프이다. 도 6c는 바이러스성 간질환 그룹 내 세부 간질환 분류를 위해 개발된 머신 러닝 기반 분류 모델 정확도 평가 결과의 ROC (Receiver operating characteristic) 그래프이다.

도 7은 본 발명의 머신 러닝 기법을 활용한 미생물 균총으로부터 간질환 군의 감별법의 분석 방법을 설명하기 위한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0017] 이하, 첨부한 도면을 참고로 하여, 본 발명의 이해를 돕기 위하여 바람직한 실시예를 제시한다. 그러나 하기의 실시예는 본 발명을 보다 쉽게 이해하기 위하여 제공되는 것일 뿐, 하기 실시예에 의해 본 발명의 내용이 한정되는 것은 아니다.

[0018] 본 발명을 상세하게 설명하기 전에, 본 명세서에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 제한되어 해석되어서는 아니 되며, 본 발명의 발명자가 자신의 발명을 가장 최선의 방법으로 설명하기 위해서 각 용어의 정의를 적절하게 활용할 수 있다.

[0019] 더 나아가 이들 용어의 개념은 본 발명의 기술적인 정의에 부합하는 의미로 해석되어야 한다.

[0020] 즉, 본 명세서에서 사용된 용어는 본 발명의 바람직한 실시예를 정의하기 위해서 활용된 것일 뿐이고, 본 발명의 내용만으로 제한하려는 의도로 사용된 것이 아니다. 이들 용어는 본 발명의 여러 가지 확장성을 고려하여 정의된 용어이다.

[0021] 또한, 이하에서, 본 발명을 설명함에 있어서, 본 발명의 핵심을 불필요하게 흐릴 수 있다고 판단되는 기술, 예를 들어, 종래 기술을 포함하는 공지 기술에 대한 상세한 설명은 생략될 수 있다.

[0022] 본 명세서에서 용어 "진단"이란, 특정 질환이 발병할 가능성이 있는지, 특정 질환이 발병할 가능성이 상대적으로 높은지, 또는 특정 질환이 이미 발병하였는지 여부를 판별하는 것을 의미한다. 본 발명의 방법은 임의의 특정 환자에 대한 특정 질환 발병 위험도가 높은 환자로써 특별하고 적절한 관리를 통하여 발병 시기를 늦추거나 발병하지 않도록 하는데 사용할 수 있다. 또한, 본 발명의 방법은 특정 질환을 조기에 진단하여 가장 적절한 치료방식을 선택함으로써 치료를 결정하기 위해 임상적으로 사용될 수 있다.

[0023] 본 명세서에서 “개체의 생물학적 시료” 또는 “검사 대상 객체로부터 수득한 생물학적 시료”는 혈액, 소변, 대변, 분변일 수 있고, 상기 혈액은 바람직하게는 전혈, 혈청, 혈장, 또는 혈액 단핵구일 수 있으나, 이에 제한되는 것은 아니다.

[0024] **실시예 1: 머신 러닝을 기반으로 한 미생물 균총 분석**

[0025] 도 1a는, 환자로부터 획득한 시료 샘플을 알코올성 간염, 알코올성 간경화, 알코올성 간암, 비알코올성 간염, 바이러스성 간염, 바이러스성 간경화 질환으로 구분하는 과정을 나타낸 모식도이다. 도 1b는, 머신 러닝 기반으로 간질환 그룹 간 감별하는 방법을 나타낸 모식도이다.

[0026] 한림대학교 의과대학으로부터 환자 대변 샘플 950개를 제공받았다. 분류 기준에 적합하지 않아 제외된 220개의 샘플을 제외하고, 730개의 샘플을 대조군, 알코올성 간질환, 비알코올성 간질환 및 바이러스성 간질환으로 분류한다. 구별 기준은 다음과 같다:

[0027] 알코올성 간질환 그룹 (Group): [Aspartate aminotransferase (AST) 또는 Alanine aminotransferase (ALT). ≥ 50 IU/L 및 AST/ALT >2] 그리고 excessive alcohol consumption (male >60 및 female >40 g/wk)

[0028] 비알코올성 간질환 그룹 (Group): Aspartate aminotransferase (AST) 또는 Alanine aminotransferase (ALT) ≥ 50 IU/L.

[0029] 바이러스성 간질환 그룹 (Group): Hepatitis B 또는 C virus 보균자.

[0030] 상기 기준에 따라 구별한 결과, 알코올성 간질환 그룹, 비알코올성 간질환 그룹, 바이러스성 간질환 그룹 및 대조군 샘플의 개수는, 각각 263, 201, 115, 151개였다.

[0031] 이후, 각 샘플의 16S Ribosomal RNA sequencing 정보를 획득하였다. 구체적으로는, ㈜ 천랩에서 Illumina MiSeq platform using reagent kit V3 in PE 250 bp mode 로 16S Ribosomal RNA sequencing을 수행하였다.

[0032] 이어서, Quality filtering을 위한 QIIME2 분석을 수행하였다. Quality filtering을 위한 기준은 다음과 같다:

[0033] (1) reads with a length of <150 bp, (2) reads with an average Phred score of <20, (3) reads containing ambiguous bases, and (4) reads containing mononucleotide repeats of >8 bp.

[0034] 이어서, Feature Table & Phylogenic Tree (QIIME2 분석) 작성을 수행하였다.

[0035] 이어서, OTU table 작성 및 Taxonomic analysis를 수행하였다. 구체적으로는, SILVA reference database의 VSEARCH sequence를 활용하여 OUT table 작성 및 Taxonomic analyhsis를 수행하였다.

[0036] 이어서, feature reduction을 수행하였다. 구체적으로는, feature reduction은 독립성분분석 (ICA), 주성분 분석 (PCA), 그리고 랜덤 투영(RP)으로 구현되었다.

[0037] **실시예 2: 알코올성 간질환 진단**

[0038] 상기 실시예 1에서 수행된, 알코올성 간질환 분류를 아래에 나타낸다. 도 2a는 알코올성 간질환 그룹에서 알코올성 간질환 대조군, 간염, 간경화, 그리고 간암에 대한 미생물 균총 알파 다양성을 표현한 그래프이다.

[0039] 도 2b는 알코올성 간질환 그룹에서 알코올성 간질환 대조군, 간염, 간경화, 그리고 간암에 대한 미생물 균총 베타 다양성을 표현한 그래프이다.

[0040] 도 2c는 알코올성 간질환 그룹에서 알코올성 간질환 대조군, 간염, 간경화, 그리고 간암 그룹 간 Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria, 그리고 Verrucomicrobia 의 분포도를 6가지 색상별로 시각화 하였다.

[0041] 도 2d는 LDA (linear discriminant analysis) score 2 이상을 가진 분류학적 분기도 (Cladogram)이며, 문 (Phylum) 이하 33 가지 균총을 트리(tree) 화 하였다.

[0042] 도 2e는 각 4 개 그룹별 (알코올성 간질환 대조군, 간염, 간경화, 간암) LDA effect size를 나타내며, 4 개의 색상으로 시각화하였다. 그래프가 긴 로드 일 수록 LDA 지수가 높은 것이다. 즉 높은 지수일수록 그룹간의 통계적 차이를 특성화 하는 바이오마커로 활용될 수 있음을 나타낸다.

[0043] 간염 그룹에서는 Bacteroidaceae, 그리고 Bacteroides 가 가장 높은 LDA 수치 약 5를 나타냈으며, 간경화 그룹에서는 Anaerosinus, 그리고 Glycerini 가 가장 높은 LDA 수치 약 4.5를 나타냈으며, 간암 그룹에서는 Proteobacteria, Gammaproteobacteria, Enterobacteiales, 그리고 Enterobacteriales 가 가장 LDA 높은 수치 약 5를 나타냈다. 상기 언급된 각 균총들은 알코올성 간질환 그룹별 바이오 마커로 활용 가능성을 보여 주고 있다.

[0044] 도 2f는 각 4 개 그룹별 속 (Genus) 과 종 (Species) 히트 맵 매트릭스 (Matrix)를 표현한 다이어그램이며, 다른 3 그룹 간 비교 시 알코올성 간질환 대조군에서는 Roseburia, Bacteriodes, 그리고 Alistipes 가 높은 분포도를 보였으며, 간염 그룹에서는 Ruminococcus, 간경화 그룹에서는 Streptococcus, Dorea, Bacteroides, 그리고 Anaerohabdus, 간암 그룹에서는 Bifidobacterium 이 상대적으로 높은 분포도를 보였다.

[0045] 상기 결과로부터, Proteobacteria, Betaproteobacteria, Gammaproteobacteria, Enterobacteriales, Enterobacteriaceae, Streptococcaceae, Paraprevotellaceae, Lactobacillus, Streptococcus, Bacillus (p-value < 0.05), Bifidum 및/또는 Hamsteri 이 알코올성 간질환 대조군 보다 많이 분포하였을 경우 알코올성 간암 그룹으로 판정할 수 있음을 알 수 있다.

[0046] 또한, 상기 결과로부터, Firmicutes, Coriobacteriia, Coriobacteriales, Aerococcaceae, Collinsella, Granulicatella, Anaerosinus, Clostridium, Balaenopterae, Paraplantarum, Bolteae, Glycerini, 및/또는 spiroforme 이 알코올성 간질환 대조군 보다 많이 분포하였을 경우 알코올성 간경화 그룹 (Alcoholic cirrhosis group)으로 판정할 수 있음을 알 수 있다.

[0047] 또한, 상기 결과로부터, Erysipelotrichim, Erysipelotrichales, Bacteroidaceae, Butyricimonas, Bacteroides, 및/또는 equi 이 알코올성 간질환 대조군 보다 많이 분포하였을 경우 알코올성 간염 그룹 (Alcoholic hepatitis group)으로 판정할 수 있음을 알 수 있다.

[0048] 또한, 상기 결과로부터, Deltaproteobacteria, Bacteroidales, Clostridiales, Burkholderiales, Odoribacteraceae, S24_7, Lachnospiraceae, Ruminococcaceae, Erysipelotrichaceae, Desulfovibrionaceae, Odoribacter, Paraprevotella, Alistipes, Clostridium, Roseburia, Faecalibacterium, Gemmiger, Finegoldii, Putredinis, Inulinivorans, Lactaris, Islandicum 및/또는 Formicilis 가 다른 3개의 질환군보다 많이 분포하였을 경우 알코올성 간질환 대조군으로 판정할 수 있음을 알 수 있다.

[0049] **실시예 3: 비알코올성 간질환 진단**

[0050] 상기 실시예 1에서 수행된, 비알코올성 간질환 분류를 아래에 나타낸다. 도 3a는 비알코올성 간질환 그룹에서 비알코올성 간질환 대조군, 그리고 간염에 대한 미생물 균총 알파 다양성을 표현한 그래프이다.

[0051] 도 3b는 비알코올성 간질환 그룹에서 비알코올성 간질환 대조군, 그리고 간염에 대한 미생물 균총 베타 다양성을 표현한 그래프이다.

[0052] 도 3c는 비알코올성 간질환 그룹에서 비알코올성 간질환 대조군 그리고 간염 그룹 간 *Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria*, 그리고 *Verrucomicrobia* 의 분포도를 6가지 색상별로 시각화 하였다.

[0053] 도 3d는 LDA (linear discriminant analysis) score 2 이상을 가진 분류학적 분기도(Cladogram), 문 (Phylum) 이하 22 가지 균총을 트리(tree) 화 하였다.

[0054] 도 3e는 각 2 개 그룹별 LDA effect size를 나타내며, 2 개의 색상으로 시각화 하였다. 그래프가 긴 로드 일 수록 LDA 지수가 높은 것이다. 즉 높은 지수일수록 그룹 간의 통계적 차이를 특성화 하는 바이오마커로 활용될 수 있음을 나타낸다. 비알코올성 간질환 대조군에서는 *Bacteroidia*, *Bacteroidaceae*, 그리고 *Bacteroides* 가 LDA 지수가 가장 높은 수치인 4.8, 4.7 그리고 4.6을 나타냈다. 특이적으로, 간염 그룹에서는 *Bacilli, Mobilis*, 그리고 *Succinispira* 가 LDA 지수가 가장 낮은 수치 값인 -4.6, 그리고 -4.5인 것으로 나타났다. 이에 따라 상기 언급된 각 균총들은 비알코올성 간질환 그룹별 바이오 마커로 활용 가능성을 보여 주고 있다.

[0055] 도 3f는 각 2 개 그룹별 속 (Genus) 과 종 (Species) 히트 맵 매트릭스 (Matrix)를 표현한 다이어그램이며, 다른 2 그룹 간 비교 시 알코올성 간질환 비알코올성 간질환 대조군에서는 *Ruminococcus*, *Roseburia*, 그리고 *Macellibacteroides* 가 높은 분포도를 보였으며, 간염 그룹에서는 *Streptococcus* 가 상대적으로 높은 분포도로 나타났다.

[0056] 상기 결과로부터, 문 (Phylum) 수준에서 Actinobacteria, Firmicutes 및/또는 TM7, 강 (Class) 수준에서 Actinobacteria, Coriobacteriia, Bacilli, Erysipelotrichi, 및/또는 TM7_3, 목 (Order) 수준에서 Bifidobacteriales, Coriobacteriales, 및/또는 Erysipelotrichales, 과 (Family) 수준에서 Bifidobacteriaceae, Aerococcaceae, Lactobacillaceae, 및/또는 Streptococcaceae, 속 (Genus) 수준에서 Collinsella, Granulicatella, Weissella, Streptococcus, Butyricicoccus, Subdoligranulum, Anaerosinus, Succinispira, Catenibacterium, 및/또는 Holdemania, 그리고 종 (Species) 수준에서 Coprophilus,

Indistinctus, Balaenopterae, Hellenica, Equi, Bolteae, Pullicaecorum, Variabile, Glycerini, Mobilis, Biforme, 및/또는 Mitsuokai 가 비알코올성 간염 대조군 보다 많이 분포하였을 경우 비알코올성 간염 그룹(Non-Alcoholic hepatitis group) 으로 판정할 수 있음을 알 수 있다.

[0057] 또한 상기 결과로부터, 강 (Class) 수준에서 Bacteroidia 및/또는 Betaproteobacteria, 목 (Order) 수준에서 Desulfovibrionales, 과 (Family) 수준에서 Bacteroidaceae, 속 (Genus) 수준에서 Bacteroides, 그리고 종 (Species) 수준에서 Hamsteri, Clostridium, 및/또는 D168 이 비알코올성 간염 그룹 (Non-Alcoholic hepatitis group) 보다 낮게 분포하였을 경우 비알코올성 간질환 대조군으로 판정할 수 있음을 알 수 있다.

[0058] **실시예 4: 바이러스성 간질환 진단**

[0059] 상기 실시예 1에서 수행된, 바이러스성 간질환 분류를 아래에 나타낸다. 도 4a는 바이러스성 간질환 그룹에서 바이러스성 간질환 대조군, 그리고 간염에 대한 미생물 균총 알파 다양성을 표현한 그래프이다.

[0060] 도 4b는 바이러스성 간질환 그룹에서 바이러스성 간질환 대조군, 그리고 간염에 대한 미생물 균총 베타 다양성을 표현한 그래프이다.

[0061] 도 4c는 바이러스성 간질환 그룹에서 바이러스성 간질환 대조군, 간염, 그리고 간경화 그룹 간 Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria, 그리고 Verrucomicrobia 의 분포도를 6가지 색상별로 시각화 하였다.

[0062] 도 4d는 LDA (linear discriminant analysis) score 2 이상을 가진 분류학적 분기도(Cladogram)이며, 문(Phylum) 이하 28 가지 균총을 트리(tree)화 하였다.

[0063] 도 4e는 각 3 개 그룹별 LDA effect size를 나타내며, 각 3 개 그룹별 LDA effect size를 나타내며, 3 개의 색상으로 시각화 하였다. 그래프가 긴 로드 일 수록 LDA 지수가 높은 것이다. 즉 높은 지수일수록 그룹 간의 통계적 차이를 특성화 하는 바이오 마커로 활용될 수 있음을 나타낸다. 특이적으로, 바이러스성 간질환 대조군에서는 Bacillales 가 LDA 수치가 가장 높은 약 3.8 로 나타났으며, 간염 그룹에서는 Selenomonas, 그리고 Bovis 가 가장 높은 LDA 수치 4.5 로 나타났다. 반면에 간경화 그룹에서는 Lactobacillales 가 가장 높은 LDA 수치 4.3 으로 확인되었다. 이에 따라 상기 언급된 각 균총들은 바이러스성 간질환 그룹별 바이오 마커로 활용 가능성을 보여 주고 있다.

[0064] 도 4f는 각 3 개 그룹별 속 (Genus) 과 종 (Species) 히트 맵 매트릭스 (Matrix)를 표현한 다이어그램이며, 다른 2 개의 질환 그룹간 비교 시, 바이러스성 간질환 대조군에서는 Roseburia, Faecalibacterium, 그리고 Alistipes 가 높은 분포도로 나타났으며, 간염 그룹에서는 Succinispira, Dorea, 그리고 Anaerosinus 가 높은 분포도를 보였다. 간경화 그룹에서는 Bacteroides, 그리고 Streptococcus 가 높은 분포도로 나타났다.

[0066] 한편, 도 5는 본 개시의 일 실시 예에 따른 전자 장치의 구성 및 동작을 설명하기 위한 블록도이다. 도 6의 전자 장치(100)는, 상술한 바이오 마커를 바탕으로 훈련되어 질병을 분류할 수 있는 인공지능 모델(111)을 활용할 수 있다.

[0067] 도 5를 참조하면, 전자 장치(100)는 메모리(110) 및 프로세서(120)를 포함한다. 전자 장치(100)는 의료 진단 장치, 의료 검사 장치 등으로 구현될 수 있으며, 이밖에 적어도 하나의 소프트웨어 내지는 애플리케이션을 제공하는 서버로 구현될 수도 있다. 이 밖에도, 전자 장치(100)는 데스크탑 PC, 스마트폰, 태블릿 PC 등 다양한 단말 기기로 구현될 수 있다.

[0068] 메모리(110)는 전자 장치(100)의 구성요소들의 전반적인 동작을 제어하기 위한 운영체제(OS: Operating System) 및 전자 장치(100)의 구성요소와 관련된 적어도 하나의 인스트럭션 또는 데이터를 저장하기 위한 구성이다.

[0069] 메모리(110)는 ROM, 플래시 메모리 등의 비휘발성 메모리를 포함할 수 있으며, DRAM 등으로 구성된 휘발성 메모리를 포함할 수 있다. 또한, 메모리(110)는 하드 디스크, SSD(Solid state drive) 등을 포함할 수도 있다.

[0070] 도 5를 참조하면, 메모리(110)는 간질환을 식별/분류하기 위한 인공지능 모델(111)을 포함한다. 인공지능 모델(111)은 알코올성 간질환, 비알코올성 간질환, 바이러스성 간질환 등에 속하는 적어도 하나의 질환을 분류하기 위한 모델이다.

[0071] 인공지능 모델(111)은 특징 추출 모듈 및 간질환 분류 모듈로 구성될 수 있다.

[0072] 특징 추출 모듈은 입력 데이터의 차원을 감소하여 특징을 추출하기 위한 구성이다. 특징 추출 모듈은, PCT(Principal Component Analysis) 알고리즘, RP(Random Projection) 알고리즘, ICA(Independent Component Analysis) 알고리즘 중 적어도 하나에 해당할 수 있다. 또한, 특징 추출 모듈은 추출되는 특징 항목(ex. 바이오 마커)의 수에 따라 20개, 40개, 60개, 80개 등 다양한 수의 특징 항목을 추출하도록 설정될 수 있다. 또한, 특징 추출 모듈은 LDA(Latent Dirichlet Allocation, LDA) 방식으로 특징 항목을 추출할 수도 있다.

[0073] 구체적으로, 특징 추출 모듈은 16S Ribosomal RNA sequencing 결과 정보의 차원을 축소하여 특징을 추출할 수 있다. 이때, 특징 추출 모듈은 알코올성 간질환, 비알코올성 간질환, 바이러스성 간질환 등과 관련된 상술한 다양한 바이오 마커에 대한 데이터를 추출하도록 훈련될 수 있다.

[0074] 구체적인 예로, 인공지능 모델(111)이 비알코올성 간염을 분류하기 위한 모델인 경우를 가정한다.

[0075] 이 경우, 특징 추출 모듈은, 액티노박테리아(Actinobacteria), 피르미쿠테스(Firmicutes) 및 TM7으로 이루어진 군에서 선택된 1종 이상의 문(phylum) 미생물의 분포와 관련된 특징 정보를 추출하도록 훈련될 수 있다.

[0076] 또한, 특징 추출 모듈은, 액티노박테리아(Actinobacteria), 코리오박테리아(Coriobacteriia), 바실리(Bacilli), 에리시페로트리치(Erysipelotrichi) 및 TM7_3으로 이루어진 군에서 선택된 1종 이상의 강(class) 미생물의 분포와 관련된 특징 정보를 추출하도록 훈련될 수 있다.

[0077] 또한, 특징 추출 모듈은, 비피도박테리아레스(Bifidobacteriales), 코리오박테리아레스(Coriobacteriales) 및 에리시페로트리차레스(Erysipelotrichales)로 이루어진 군에서 선택된 1종 이상의 목(order) 미생물의 분포와 관련된 특징 정보를 추출하도록 훈련될 수 있다.

[0078] 또한, 특징 추출 모듈은, 비피도박테리아시에(Bifidobacteriaceae), 아이로콕카시에(Aerococcaceae), 락토바실라시에(Lactobacillaceae) 및 스트렙토코카시에(Streptococcaceae)로 이루어진 군에서 선택된 1종 이상의 과(family) 미생물의 분포와 관련된 특징 정보를 추출하도록 훈련될 수 있다.

[0079] 또한, 특징 추출 모듈은, 콜린셀라(Collinsella), 그라눌리카텔라(Granulicatella), 웨이셀라(Weissella), 스트렙토코쿠스(Streptococcus), 부티리시코쿠스(Butyricicoccus), 서브돌리그라눌럼(Subdoligranulum), 아네로시누스(Anaerosinus), 숙시니스피라(Succinispira), 카테니박테리움(Catenibacterium) 및 홀데마니아(Holdemania)로 이루어진 군에서 선택된 1종 이상의 속(genus) 미생물의 분포와 관련된 특징 정보를 추출하도록 훈련될 수 있다.

[0080] 또한, 특징 추출 모듈은, 코프로피루스(Coprophilus), 인디스틴크투스(Indistinctus), 발라에노프테라에(Balaenopterae), 헬레니카(Hellenica), 이콰이(Equi), 볼티에(Bolteae), 풀리카에코룸(Pullicaecorum), 바리아빌레(Variabile), 글리세리니(Glycerini), 모빌리스(Mobilis), 비포르메(Biforme) 및 미트수오카이(Mitsuokai)로 이루어진 군에서 선택된 1종 이상의 종(species) 미생물의 분포와 관련된 특징 정보를 추출하도록 훈련될 수 있다.

[0081] 간질환 분류 모듈은 간질환의 유형을 식별하기 위한 모듈이다. 간질환 분류 모듈은 특징 추출 모듈에서 추출된 특징 정보를 바탕으로 적어도 하나의 간질환을 선택할 수 있다.

[0082] 간질환 분류 모듈은 SVM(Support Vector Machine), RF(Random Forest), MLP(Multi-Layer Perceptron), CNN(Convolutional Neural Network) 중 적어도 하나의 모델에 해당할 수 있다.

[0083] SVM은 분류를 위한 기준선을 정의하는 모델로, 분류되지 않은 새로운 점이 나타나면 경계의 어느 쪽에 속하는지 확인하여 분류를 수행하는 모델이다. 이때, 하이퍼 파라미터로 지정할 수 있는 값들을 순차적으로 입력한 뒤 가장 높은 성능을 보이는 하이퍼 파라미터를 찾는 Grid Search가 사용될 수 있다. 본 모델의 구성과 관련된 하이퍼 파라미터는 C(Cost. 어느 정도의 오차를 허용할지에 대한 파라미터), kernel(매핑된 공간에서 결정 마진을 계산하기 위한 함수), gamma(결정 경계를 얼마나 유연하게 그을 것인지에 대한 파라미터) 등을 포함할 수 있으며, 다양한 하이퍼 파라미터 값의 조합에 따라 시험이 수행되었다.

[0084] RF는 복수의 의사결정나무(Decision Trees)를 훈련시켜 그 결과를 종합해 예측하는 앙상블 알고리즘이다. 본 모델의 구성과 관련된 하이퍼 파라미터는 n estimators(결정 트리의 개수), max features(최대 feature 개수), min samples split(노드를 분할하기 위한 최소한의 샘플 데이터 수) 등을 포함할 수 있으며, 역시 다양한 하이퍼 파라미터 값의 조합에 따라 시험이 수행되었다.

[0085] MLP는 퍼셉트론으로 이루어진 층이 여러 개 순차적으로 붙여진 형태의 모델이다. MLP 모델은 입력층, 은닉층,

출력층을 포함하며, 일 예로, 은닉층에 ReLu 활성화함수가 포함되고, 출력층에 sigmoid 활성화 함수가 포함될 수 있다. 본 모델의 구성과 관련된 하이퍼 파라미터는 numHiddenLayer(은닉층 수), epochs(학습 횟수), numUnits(은닉층의 유닛 수), dropout rate(오버피팅 방지를 위한 뉴런 제거 확률) 등을 포함할 수 있다. 역시 다양한 하이퍼 파라미터 값의 조합에 따라 시험이 수행되었다.

[0086] CNN은 컨볼루션 레이어, 맥스 풀링 레이어, 풀리 커넥티드 레이어 등을 포함할 수 있다. 본 모델의 구성과 관련된 하이퍼 파라미터는 epochs(학습 횟수), optimizer(손실 함수의 값을 줄이기 위한 가중치 및 학습 비율과 관련된 속성 설정), loss(손실 함수) 등을 포함할 수 있다. 역시 다양한 하이퍼 파라미터 값의 조합에 따라 시험이 수행되었다.

[0087] 프로세서(120)는 전자 장치(100)를 전반적으로 제어하기 위한 구성이다. 구체적으로, 프로세서(120)는 메모리(110)와 연결되는 한편 메모리(110)에 저장된 적어도 하나의 인스트럭션을 실행함으로써 본 개시의 다양한 실시예들에 따른 동작을 수행할 수 있다.

[0088] 프로세서(120)는 CPU, AP, DSP(Digital Signal Processor) 등과 같은 범용 프로세서, GPU, VPU(Vision Processing Unit) 등과 같은 그래픽 전용 프로세서 또는 NPU와 같은 인공지능 전용 프로세서 등을 포함할 수 있다. 인공지능 전용 프로세서는, 특정 인공지능 모델의 훈련 내지는 이용에 특화된 하드웨어 구조로 설계될 수 있다.

[0089] 프로세서(120)는 16S Ribosomal RNA sequencing 결과 정보를 인공지능 모델(111)에 입력하여 적어도 하나의 간질환(ex. 비알코올성 간염 등)에 대한 해당 여부를 식별할 수 있다.

[0090] 프로세서(120)는 다양한 유형의 간질환에 해당하는 복수의 환자의 16S Ribosomal RNA sequencing 결과 정보를 바탕으로 인공지능 모델(111)을 훈련시킬 수 있다.

[0091] 이때, 프로세서(120)는 하나 이상의 유형의 간질환과 연관성이 있는 바이오마커를 추출하도록 특징 추출 모듈을 훈련시킬 수 있으며, 바이오마커의 수치에 따라 적어도 하나의 간질환의 해당 여부를 판단하도록 간질환 분류 모듈을 훈련시킬 수 있다.

[0093] **실시예 5: 정확도 평가(인공지능 모델)**

[0094] 도 6a는 알코올성 간질환 그룹 내 대조군, 간염, 간경화, 그리고 간암 그룹으로 구성되는 4개의 그룹 중 2개 그룹의 각 조합(총 6개)에 속하는 그룹 간 분류를 위해 개발된 머신 러닝 기반 분류 모델 정확도 평가 결과의 ROC (Receiver operating characteristic) 그래프이다. 여기서 짙은색의 실선은 10-fold cross validation에 의해 시행된 10번 (연한색의 실선들)의 모델 정확도 평가 결과의 평균을 나타내며, 각 실선들은 60개의 서로 다른 hyper parameter 설정을 통해 얻은 결과들을 도시한 것이다. 각 그래프의 삽입은 10-fold cross validation에 의해 시행된 10번의 모델 정확도 평가 결과를 AUC (Area under the curve)의 값으로 나열하고 있다.

[0095] 도 6b는 비알코올성 간질환 그룹 내 대조군과 간염 간 분류를 위해 개발된 머신 러닝 기반 분류 모델 정확도 평가 결과의 ROC (Receiver operating characteristic) 그래프이다. 여기서 짙은색의 실선은 10-fold cross validation에 의해 시행된 10번 (연한색의 실선들)의 모델 정확도 평가 결과의 평균을 나타내며, 각 실선들은 60개의 서로 다른 hyper parameter 설정을 통해 얻은 결과들을 도시한 것이다. 각 그래프의 삽입은 10-fold cross validation에 의해 시행된 10번의 모델 정확도 평가 결과를 AUC (Area under the curve)의 값으로 나열하고 있다.

[0096] 도 6c는 바이러스성 간질환 그룹 내 대조군, 간염, 그리고 간경화 그룹으로 구성되는 3개의 그룹 중 2개 그룹의 각 조합 (총 3개) 속하는 그룹 간 분류를 위해 개발된 머신 러닝 기반 분류 모델 정확도 평가 결과의 ROC (Receiver operating characteristic) 그래프이다. 여기서 짙은색의 실선은 10-fold cross validation에 의해 시행된 10번 (연한색의 실선들)의 모델 정확도 평가 결과의 평균을 나타내며, 각 실선들은 60개의 서로 다른 hyper parameter 설정을 통해 얻은 결과들을 도시한 것이다. 각 그래프의 삽입은 10-fold cross validation에 의해 시행된 10번의 모델 정확도 평가 결과를 AUC (Area under the curve)의 값으로 나열하고 있다.

[0097] 한편, 이상에서 설명된 다양한 실시 예들은 서로 저촉되거나 모순되지 않는 한 두 개 이상의 실시 예가 서로 결합되어 구현될 수 있다.

[0098] 한편, 이상에서 설명된 전자 장치의 다양한 실시 예들은 소프트웨어(software), 하드웨어(hardware) 또는 이들

의 조합된 것을 이용하여 컴퓨터(computer) 또는 이와 유사한 장치로 읽을 수 있는 기록 매체 내에서 구현될 수 있다.

[0099] 하드웨어적인 구현에 의하면, 본 개시에서 설명되는 실시 예들은 ASICs(Application Specific Integrated Circuits), DSPs(digital signal processors), DSPDs(digital signal processing devices), PLDs(programmable logic devices), FPGAs(field programmable gate arrays), 프로세서(processors), 제어기(controllers), 마이크로 컨트롤러(micro-controllers), 마이크로 프로세서(microprocessors), 기타 기능 수행을 위한 전기적인 유닛(unit) 중 적어도 하나를 이용하여 구현될 수 있다.

[0100] 일부의 경우에 본 명세서에서 설명되는 실시 예들이 프로세서 자체로 구현될 수 있다. 소프트웨어적인 구현에 의하면, 본 명세서에서 설명되는 절차 및 기능과 같은 실시 예들은 별도의 소프트웨어 모듈들로 구현될 수 있다. 상술한 소프트웨어 모듈들 각각은 본 명세서에서 설명되는 하나 이상의 기능 및 작동을 수행할 수 있다.

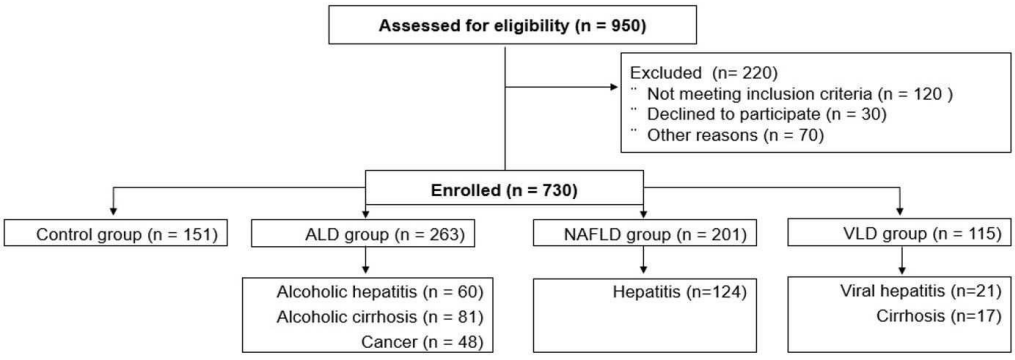
[0101] 한편, 상술한 본 개시의 다양한 실시 예들에 전자 장치에서의 처리동작을 수행하기 위한 컴퓨터 명령어(computer instructions) 또는 컴퓨터 프로그램은 비일시적 컴퓨터 판독 가능 매체(non-transitory computer-readable medium)에 저장될 수 있다. 이러한 비일시적 컴퓨터 판독 가능 매체에 저장된 컴퓨터 명령어 또는 컴퓨터 프로그램은 특정 기기의 프로세서에 의해 실행되었을 때 상술한 다양한 실시 예에 따른 전자 장치에서의 처리 동작을 상술한 특정 기기가 수행하도록 한다.

[0102] 비일시적 컴퓨터 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 비일시적 컴퓨터 판독 가능 매체의 구체적인 예로는, CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM 등이 있을 수 있다.

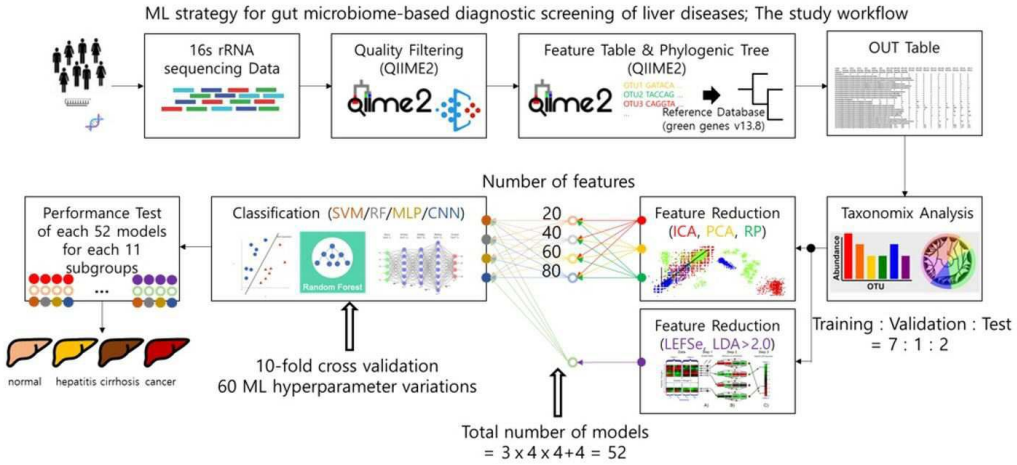
[0103] 이상에서는 본 개시의 바람직한 실시 예에 대하여 도시하고 설명하였지만, 본 개시는 상술한 특정의 실시 예에 한정되지 아니하며, 청구범위에서 청구하는 본 개시의 요지를 벗어남이 없이 당해 개시에 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 개시의 기술적 사상이나 전망으로부터 개별적으로 이해되어서는 안될 것이다.

도면

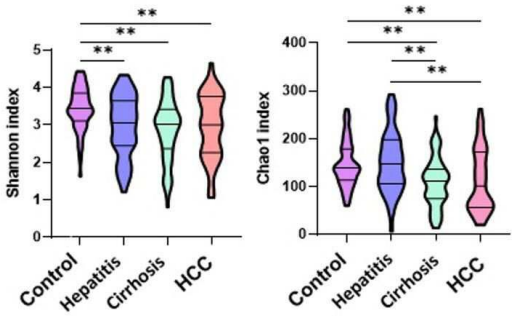
도면1a



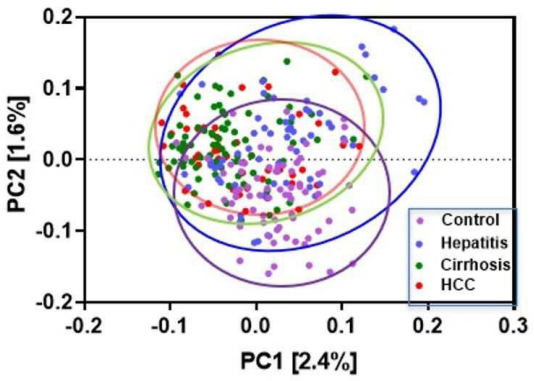
도면1b



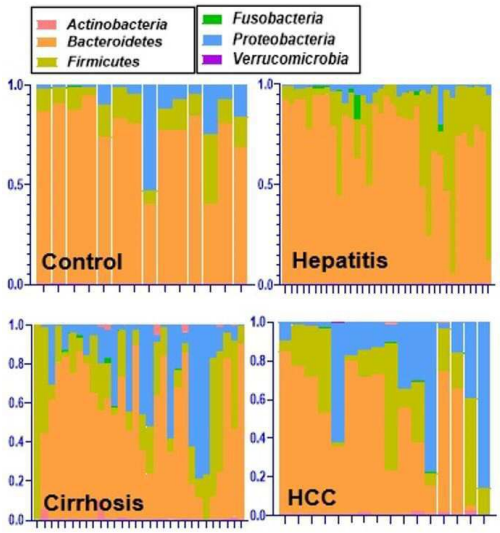
도면2a



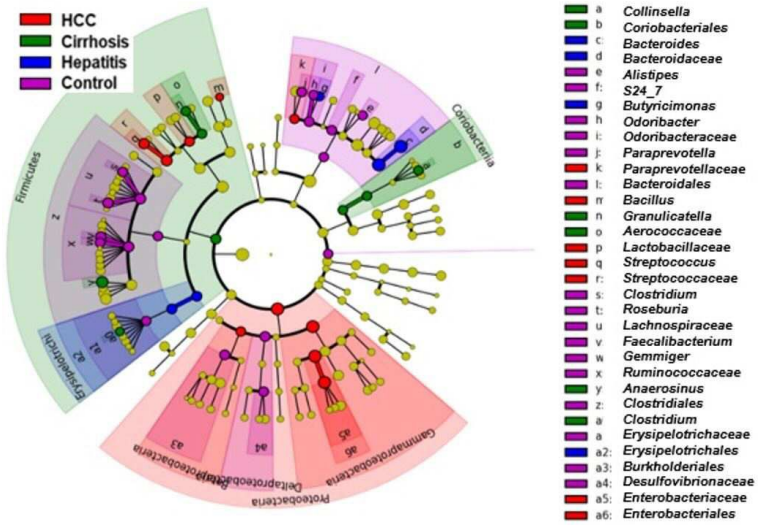
도면2b



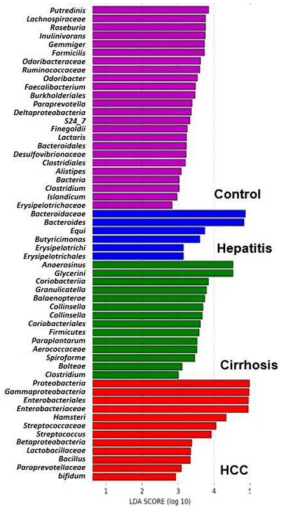
도면2c



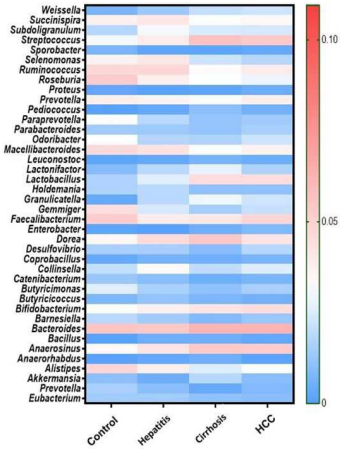
도면2d



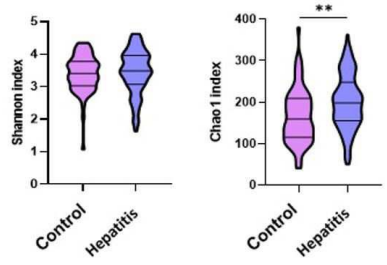
도면2e



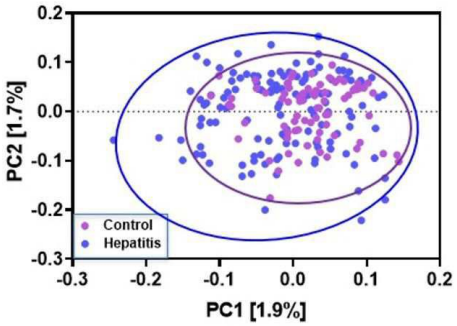
도면2f



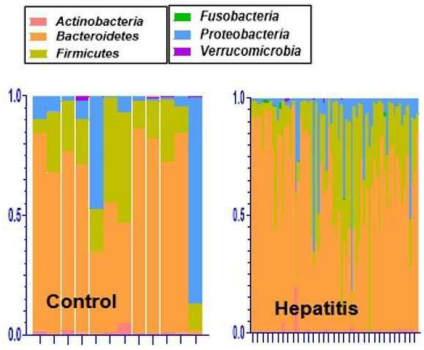
도면3a



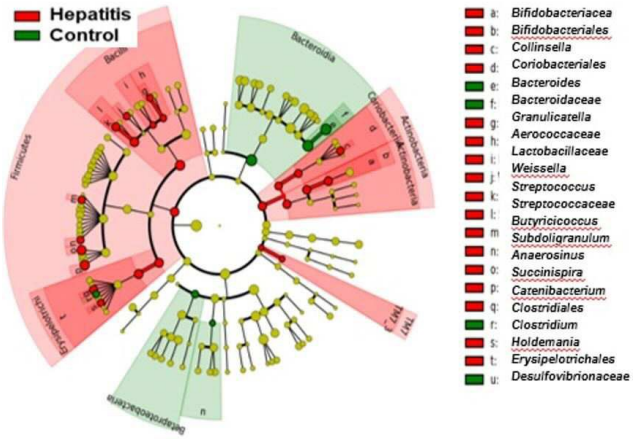
도면3b



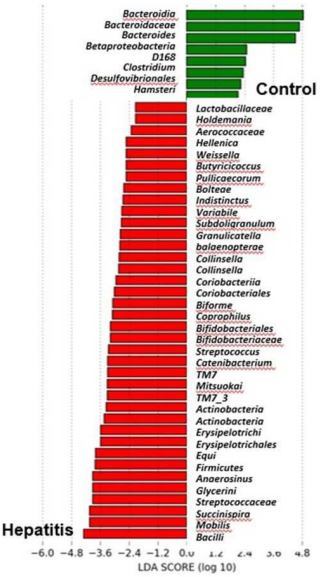
도면3c



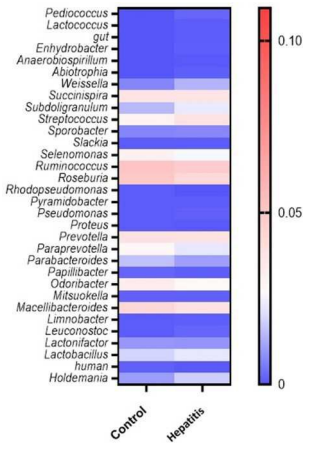
도면3d



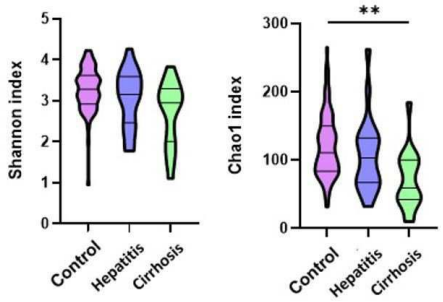
도면3e



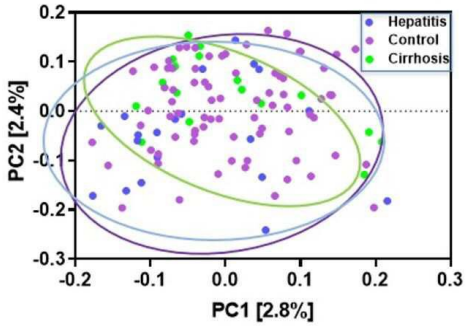
도면3f



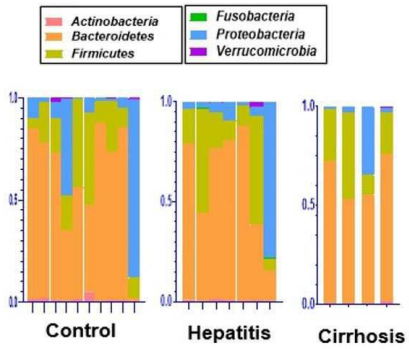
도면4a



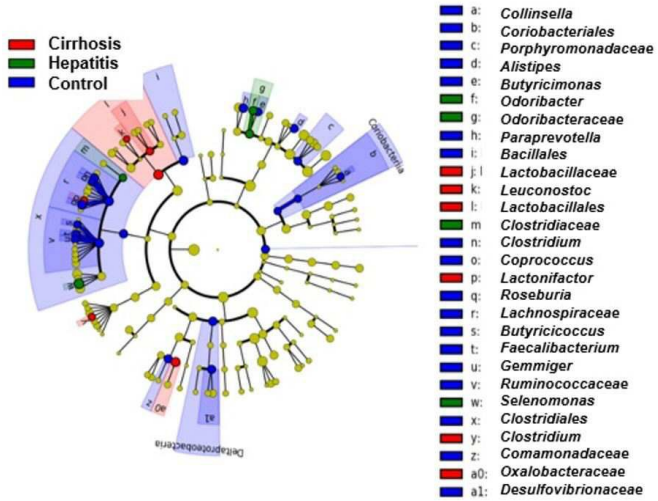
도면4b



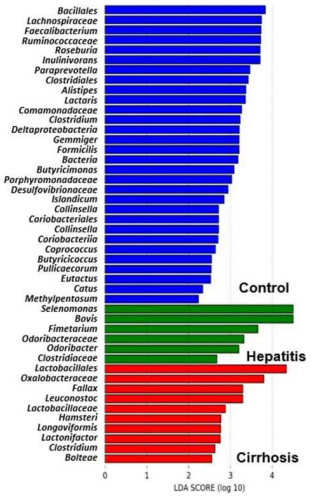
도면4c



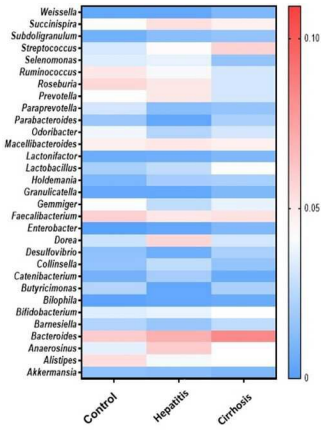
도면4d



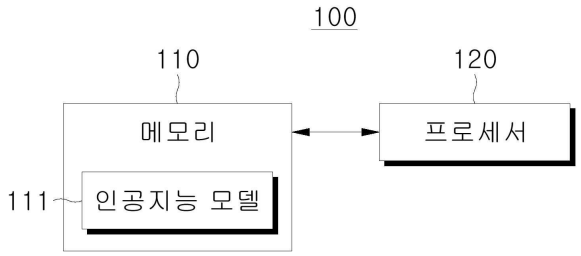
도면4e



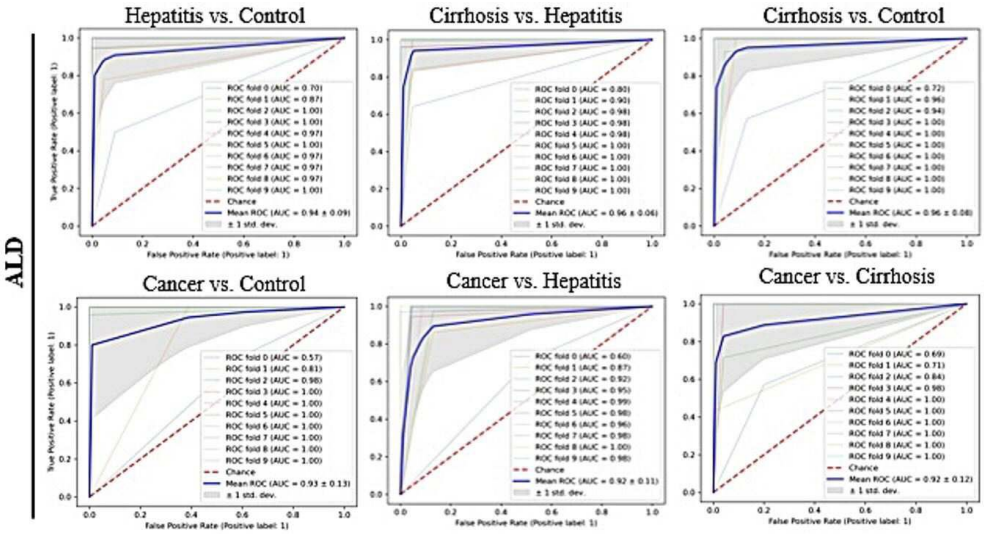
도면4f



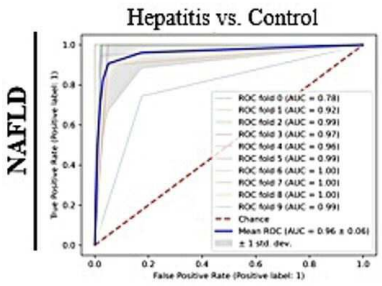
도면5



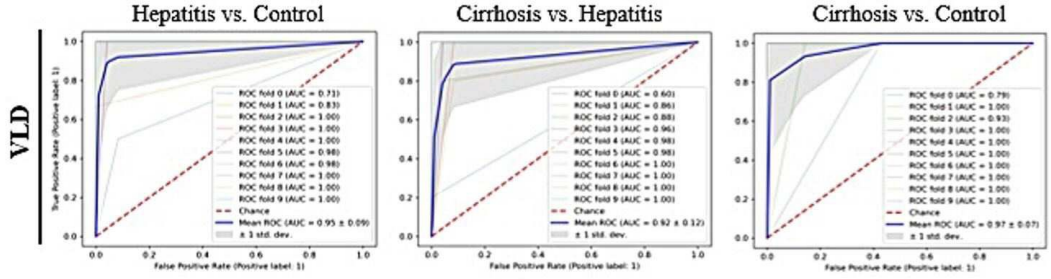
도면6a



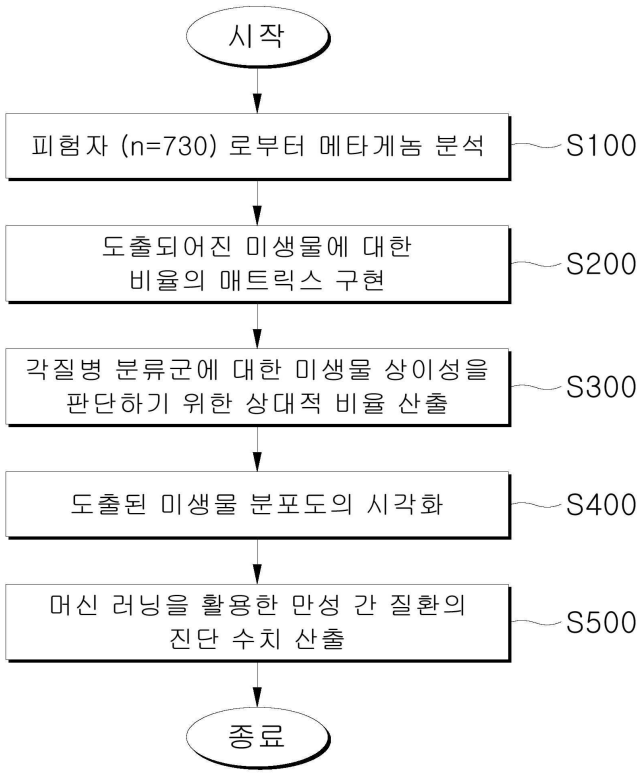
도면6b



도면6c



도면7


시작
피험자 (n=730) 로부터 메타게놈 분석
S100
도출되어진 미생물에 대한
비율의 매트릭스 구현
S200
각질병 분류군에 대한 미생물 상이성을
판단하기 위한 상대적 비율 산출
S300
도출된 미생물 분포도의 시각화
S400
머신 러닝을 활용한 만성 간 질환의
진단 수치 산출
S500
종료