(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0032225
(43) 공개일자 2025년03월07일

(51) 국제특허분류(Int. Cl.)
G16H 50/20 (2018.01) *A61B 5/00* (2021.01)
G06N 20/00 (2019.01) *G16H 10/60* (2018.01)
G16H 50/50 (2018.01) *G16H 50/70* (2018.01)
(52) CPC특허분류
G16H 50/20 (2018.01)
A61B 5/4088 (2013.01)
(21) 출원번호 10-2023-0114773
(22) 출원일자 2023년08월30일
심사청구일자 2023년08월30일

(71) 출원인
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)
(72) 발명자
김건하
서울특별시 서초구 동광로19길 51 1105호
정지향
서울특별시 마포구 양화로 26, 801호
(뒷면에 계속)
(74) 대리인
특허법인한일

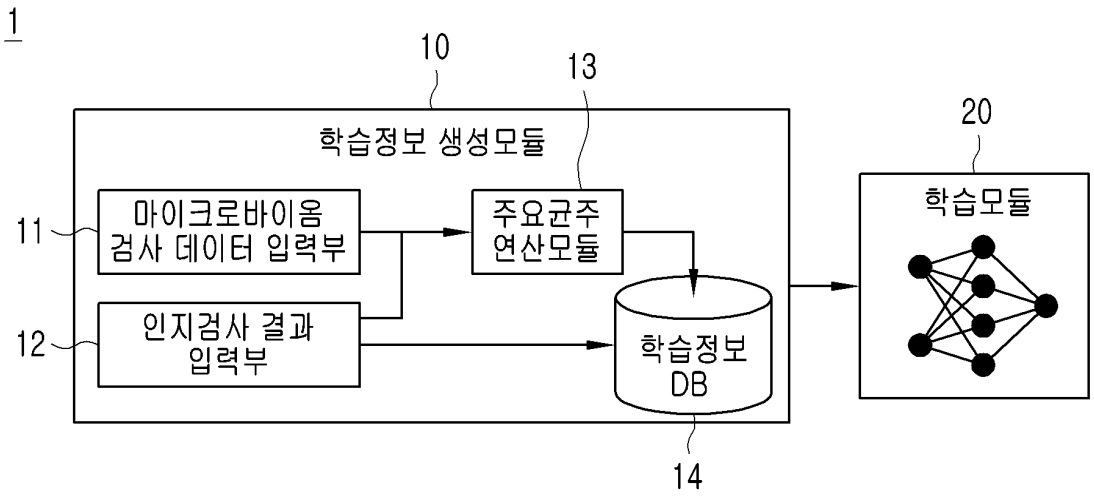
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 **마이크로바이옴 기반 슈퍼에이저 판단 방법 및 장치**

(57) 요 약

본 개시의 일 실시예에 의하면, (a1) 마이크로바이옴 검사 데이터 입력부에 입력된 마이크로바이옴 검사 데이터 및 인지검사 결과 입력부에 입력된 인지검사 결과를 이용하여, 주요균주 연산모듈이 주요균주를 선별하는 단계; (a2) 선별된 주요균주의 상대존재비(relative abundance)가 주요균주 연산모듈에 의해 확인되고, 상기 확인된 주요균주의 상대존재비 및 상기 인지검사 결과가 학습정보 데이터세트로서 학습정보 데이터베이스에 저장되는 단계; (a3) 상기 주요균주의 상대존재비를 입력자료로 하고 상기 인지검사 결과를 출력자료로 하는 학습모델이 학습모듈에 의해 학습되는 단계; (b1) 상기 학습모듈에 주요균주의 상대존재비가 입력되면, 예측 인지검사 결과가 출력되는 단계; 및 (b2) 상기 (b1) 단계에서 확인되는 상기 인지검사 결과를 이용하여 슈페어이저 여부가 확인되는 단계;를 포함하는, 방법을 제공한다.

대 표 도 - 도1



(52) CPC특허분류
G06N 20/00 (2021.08)
G16H 10/60 (2021.08)
G16H 50/50 (2018.01)
G16H 50/70 (2018.01)
(72) 발명자
김보리
서울특별시 강북구 솔샘로 159 벽산라이브파크아파트 101동 1602호

김하은
서울특별시 서대문구 증가로20가길 33(북가좌동)

이 발명을 지원한 국가연구개발사업

과제고유번호	1465039564
과제번호	HU20C0271010023
부처명	다부처
과제관리(전문)기관명	한국보건산업진흥원
연구사업명	치매극복연구개발사업(복지부,과기부)
연구과제명	뇌영상 분석을 통한 뇌회복력(Brain Resilience) 및 인지예비능 분석
기 여 율	4/5
과제수행기관명	이화여자대학교 산학협력단
연구기간	2023.04.01 ~ 2023.05.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179344
과제번호	00155966
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	인공지능융합혁신인재양성(R&D)
연구과제명	인공지능융합혁신인재양성(이화여자대학교)
기 여 율	1/5
과제수행기관명	이화여자대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명 세 서

청구범위

청구항 1

마이크로바이옴 검사 데이터를 이용하여 슈퍼에이저 여부를 판단하기 위한 방법에 있어서,

(a1) 마이크로바이옴 검사 데이터 입력부에 입력된 마이크로바이옴 검사 데이터 및 인지검사 결과 입력부에 입력된 인지검사 결과를 이용하여, 주요균주 연산모듈이 주요균주를 선별하는 단계;

(a2) 선별된 주요균주의 상대존재비(relative abundance)가 주요균주 연산모듈에 의해 확인되고, 상기 확인된 주요균주의 상대존재비 및 상기 인지검사 결과가 학습정보 데이터세트로서 학습정보 데이터베이스에 저장되는 단계;

(a3) 상기 주요균주의 상대존재비를 입력자료로 하고 상기 인지검사 결과를 출력자료로 하는 학습모델이 학습모듈에 의해 학습되는 단계;

(b1) 상기 학습모듈에 주요균주의 상대존재비가 입력되면, 예측 인지검사 결과가 출력되는 단계; 및

(b2) 상기 (b1) 단계에서 확인되는 상기 인지검사 결과를 이용하여 슈페어이저 여부가 확인되는 단계;를 포함하는,

방법.

청구항 2

제1항에 있어서,

상기 (a2) 단계에서, 상기 학습정보 데이터베이스에는 신상 데이터 및 신체 데이터가 더 저장되고,

상기 (a3) 단계에서, 상기 신상 데이터 및 상기 신체 데이터가 입력자료로 더 이용되고,

상기 (b1) 단계에서, 상기 학습모듈에 신상 데이터 및 신체 데이터가 더 입력되면, 인지검사 결과가 출력되는,

방법.

청구항 3

제1항에 있어서,

상기 (a2) 단계에서, 상기 주요균주 연산모듈은 Mann-Whitney U test 및 POINT Biserial Correlation 방법을 이용하여 복수의 주요균주를 선별하는,

방법.

청구항 4

제3항에 있어서,

상기 (a1) 단계에서 선별된 주요균주의 개수는 16개인,

방법.

청구항 5

제4항에 있어서,

상기 (a1) 단계에서 선별된 주요균주는 AC001115_g, Eisenbergiella, PAC001138_g, Sellimonas, PAC001236_g, Romboutsia, Eubacterium_g8, Subdoligranulum, Lactobacillaceae, Leuconostoc, Odoribacteraceae, Genus Coprococcus, Ruminococcus_g4, Actinobacteria_c, Rothia 및 Coprococcus인,

방법.

청구항 6

제1항에 있어서,

상기 (a3) 단계에서, 상기 학습모델은 Random Forest, Gradient Boosting, CatBoost, SVM(Support Vector Machine) 및 MLP(Multi Layer Perceptron) 중 어느 하나인,

방법.

청구항 7

제1항 내지 제6항 중 어느 한 항에 따른 방법이 수행되도록, 컴퓨터로 판독가능한 저장매체에 저장되는, 프로그램.

청구항 8

마이크로바이옴 검사 데이터를 이용하여 슈퍼에이저 여부를 판단하기 위한 장치에 있어서,

학습정보 생성모듈로서,

마이크로바이옴 검사 데이터가 입력되는 마이크로바이옴 검사 데이터 입력부;

인지검사 결과가 입력되는 인지검사 결과 입력부;

상기 마이크로바이옴 검사 데이터 및 상기 인지검사 결과를 이용하여 주요균주를 선별하고, 선별된 주요균주의 상대존재비를 확인하는 주요균주 연산모듈; 및

상기 주요균주의 상대존재비 및 상기 인지검사 결과를 학습정보 데이터세트로서 저장하는 학습정보 데이터베이스를 포함하는 학습정보 생성모듈; 및

상기 주요균주의 상대존재비를 입력자료로 하고, 상기 인지검사 결과를 출력자료로 하는 학습 모델을 학습시키도록 구성된 학습모듈;을 포함하고,

상기 학습모듈에 주요균주의 상대존재비가 입력되면, 예측 인지검사 결과가 출력되고,

상기 예측 인지검사 결과로부터 슈퍼에이저 여부가 확인될 수 있는,

장치.

청구항 9

제8항에 있어서,

상기 학습정보 생성모듈은,

연령, 성별 및 교육수준을 포함하는 신상 데이터가 입력되는 신상 데이터 입력부; 및

체질량지수(BMI)를 포함하는 신체 데이터가 입력되는 신체 데이터 입력부를 더 포함하고,

상기 학습정보 생성모듈은 상기 신상 데이터 및 상기 신체 데이터를 더 저장하고,

상기 학습모듈은 상기 신상 데이터 및 상기 신체 데이터를 입력자료로 더 이용하는,

장치.

발명의 설명

기 술 분 야

[0001] 본 개시는 슈퍼에이저 판단 방법 및 장치에 관한 것으로, 구체적으로는 마이크로바이옴에 기반하여, 슈퍼에이저에 해당하는지 여부를 판단하기 위한 방법 및 장치에 관한 것이다.

배 경 기 술

[0003] 이 부분에 기술된 내용은 단순히 본 개시에 대한 배경정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0004] 노인 인구가 급격히 증가하면서 노인의 건강관리가 중요한 사회적 관심사로 떠오르고 있으며, 특히 뇌 건강과 관련된 노년기 인지기능 변화에 대한 연구가 활발히 이루어지고 있다. 노년기 인지기능은 연령의 증가에 따라 기능의 손상 혹은 저하가 일반적이라고 하더라도 이러한 변화는 고정된 것이 아니라 여러가지 요인에 의해서 다르게 나타날 수 있는 개인적 속성이다.

[0005] 인지기능을 측정하는 인지기능 테스트는 주로 경도 인지 장애나 치매와 같은 질환을 판별하는 수단으로 활용되며, 서류를 이용하거나 서류를 기반으로 한 컴퓨터 프로그램으로 수행되고 있다. 그러나 이와 같은 서류기반의 인지기능 테스트는 장시간이 소요되어 노년층에게 부담이 되며, 문자를 읽는 것이 불편한 노년층의 경우 활용하기 어려운 문제점이 있다. 또한, 동일한 테스트를 다수 회 수행할 수 없어 인지기능 변화에 대한 획일적인 추적 관찰이 불가능한 문제점이 있었다.

[0006] 한편, 기존의 인지기능 관련 연구는 주로 알츠하이머, 파킨슨, 당뇨, 고혈압 등 질환이 있는 환자를 대상으로 진행되어 왔으나, 최근 해당 연령에 비하여 강인한 인지능력을 갖는 노인을 의미하는 슈퍼에이저(Super-ager)와 관련된 연구들이 주목받고 있다. 슈퍼에이저는 고령이지만 나이에 비해 월등하게 우수한 인지능력을 가짐으로 인해 신체적, 정신적으로 더욱 건강한 노후 생활을 영위한다. 따라서, 이러한 슈퍼에이저는 노인의 인지기능을 평가하는 중요한 하나의 척도가 될 수 있다.

[0007] 슈퍼에이저는 일반 노인에 비하여 뇌의 내측두엽 및 아세틸콜린 분해와 연관된 뇌영역 부피가 크다는 점에서, 오히려 신경퇴행(neurodegeneration)에 저항성 있는(resistant) 그룹으로 알려져 있다. 즉, 슈퍼에이저 그룹은 신경퇴행 그룹과는 대척점에 있는 그룹임에도, 슈퍼에이저 여부를 판단하기 위해서는 기존의 신경퇴행 여부를 검사하기 위한 인지기능 검사를 이용하여야 하는바, 슈퍼에이저에 최적화된 검사 방법 등이 미비한 실정이다.

발명의 내용

해결하려는 과제

[0009] 이에, 본 개시는 인지능력, 특히 노인 중 우수한 인지능력을 갖는 슈퍼에이저를 판단하기 위한 방법 및 장치를 제공하는 데 목적이 있다.

[0010] 또한, 본 개시는 마이크로바이옴 검사 데이터 외에도, 슈퍼에이저 여부에 영향을 미칠 수 있는 여러 항목들을 확인할 수 있는 방법 및 장치를 제공하는 데 목적이 있다.

[0011] 본 발명이 해결하고자 하는 과제들은 이상에서 언급한 과제들로 제한되지 않으며, 언급되지 않은 또 다른 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0013] 본 개시의 일 실시예에 의하면, 마이크로바이옴 검사 데이터를 이용하여 슈퍼에이저 여부를 판단하기 위한 방법에 있어서, (a1) 마이크로바이옴 검사 데이터 입력부에 입력된 마이크로바이옴 검사 데이터 및 인지검사 결과

입력부에 입력된 인지검사 결과를 이용하여, 주요균주 연산모듈이 주요균주를 선별하는 단계; (a2) 선별된 주요균주의 상대존재비(relative abundance)가 주요균주 연산모듈에 의해 확인되고, 상기 확인된 주요균주의 상대존재비 및 상기 인지검사 결과가 학습정보 데이터세트로서 학습정보 데이터베이스에 저장되는 단계; (a3) 상기 주요균주의 상대존재비를 입력자료로 하고 상기 인지검사 결과를 출력자료로 하는 학습모델이 학습모듈에 의해 학습되는 단계; (b1) 상기 학습모듈에 주요균주의 상대존재비가 입력되면, 예측 인지검사 결과가 출력되는 단계; 및 (b2) 상기 (b1) 단계에서 확인되는 상기 인지검사 결과를 이용하여 슈페어이저 여부가 확인되는 단계;를 포함하는, 방법을 제공한다.

[0014] 또한, 바람직하게는, 본 개시의 다른 실시예에 의하면, 상기 (a2) 단계에서, 상기 학습정보 데이터베이스에는 신상 데이터 및 신체 데이터가 더 저장되고, 상기 (a3) 단계에서, 상기 신상 데이터 및 상기 신체 데이터가 입력자료로 더 이용되고, 상기 (b) 단계에서, 상기 학습모듈에 신상 데이터 및 신체 데이터가 더 입력되면, 인지검사 결과가 출력된다.

[0015] 또한, 바람직하게는, 본 개시의 일 실시예에 의한 상기 (a2) 단계에서, 상기 주요균주 연산모듈은 Mann-Whitney U test 및 POINT Biserial Correlation 방법을 이용하여 복수의 주요균주를 선별한다.

[0016] 또한, 바람직하게는, 본 개시의 일 실시예에 의한 상기 (a1) 단계에서 선별된 주요균주의 개수는 16개이다.

[0017] 또한, 바람직하게는, 본 개시의 일 실시예에 의한 상기 (a1) 단계에서 선별된 주요균주는 ACO01115_g, Eisenbergiella, PAC001138_g, Sellimonas, PAC001236_g, Romboutsia, Eubacterium_g8, Subdoligranulum, Lactobacillaceae, Leuconostoc, Odoribacteraceae, Genus Coprococcus, Ruminococcus_g4, Actinobacteria_c, Rothia 및 Coprococcus이다.

[0018] 또한, 바람직하게는, 본 개시의 일 실시예에 의한 상기 (a3) 단계에서, 상기 학습모델은 Random Forest, Gradient Boosting, CatBoost, SVM(Support Vector Machine) 및 MLP(Multi Layer Perceptron) 중 어느 하나이다.

[0019] 또한, 바람직하게는, 본 개시의 실시예들에 의한 방법이 수행되도록, 컴퓨터로 판독가능한 저장매체에 저장되는, 프로그램이 제공된다.

[0020] 본 개시의 일 실시예에 의하면, 마이크로바이옴 검사 데이터를 이용하여 슈퍼에이저 여부를 판단하기 위한 장치에 있어서, 학습정보 생성모듈로서, 마이크로바이옴 검사 데이터가 입력되는 마이크로바이옴 검사 데이터 입력부; 인지검사 결과가 입력되는 인지검사 결과 입력부; 상기 마이크로바이옴 검사 데이터 및 상기 인지검사 결과를 이용하여 주요균주를 선별하고, 선별된 주요균주의 상대존재비를 확인하는 주요균주 연산모듈; 및 상기 주요균주의 상대존재비 및 상기 인지검사 결과를 학습정보 데이터세트로서 저장하는 학습정보 데이터베이스를 포함하는 학습정보 생성모듈; 및 상기 주요균주의 상대존재비를 입력자료로 하고, 상기 인지검사 결과를 출력자료로 하는 학습 모델을 학습시키도록 구성된 학습모듈;을 포함하고, 상기 학습모듈에 주요균주의 상대존재비가 입력되면, 예측 인지검사 결과가 출력되고, 상기 예측 인지검사 결과로부터 슈퍼에이저 여부가 확인될 수 있는, 장치를 제공한다.

[0021] 또한, 본 개시의 다른 실시예에 의하면, 상기 학습정보 생성모듈은, 연령, 성별 및 교육수준을 포함하는 신상 데이터가 입력되는 신상 데이터 입력부; 및 체질량지수(BMI)를 포함하는 신체 데이터가 입력되는 신체 데이터 입력부를 더 포함하고, 상기 학습정보 생성모듈은 상기 신상 데이터 및 상기 신체 데이터를 더 저장하고, 상기 학습모듈은 상기 신상 데이터 및 상기 신체 데이터를 입력자료로 더 이용하는, 장치를 제공한다.

발명의 효과

[0023] 이상에서 설명한 바와 같이 본 개시의 실시예에 의하면, 피험자의 마이크로바이옴 검사 데이터를 학습모듈에 입력하기만 하여도, 피험자의 슈퍼에이저 해당 여부가 확인될 수 있어, 인지능력 평가에 대한 사용자들의 접근성이 높아진다는 효과가 있다.

[0024] 또한, 본 개시의 실시예들에 의하면, 마이크로바이옴 검사 데이터를 기반으로 슈퍼에이저 여부가 평가됨으로써, 피험자가 인지능력 검사 시험지에 직접 답변을 체크한 경우보다 더 일관되며, 객관적인 결과가 제공된다는 효과가 있다.

[0025] 또한, 기존 인지검사 비용에 비해 월등히 낮은 비용으로 검사가 진행될 수 있고, 기존 인지검사가 장시간 검사

시간을 요구하는 데에 반해, 피험자는 채변 검사에만 참여하면 되므로, 피험자가 경제적, 시간적으로 부담을 적게 느낄 수 있다는 효과가 있다.

[0026] 또한, 매번 다른 종류의 시험지가 제공될 필요가 없고, 마이크로바이옴 데이터만 취득되면 되기에, 시간에 따른 추적관찰이 가능한바, 지속적인 건강관리에 도움이 된다는 효과가 있다.

[0027] 또한, 본 개시의 다른 실시예에 따르면, 신상 데이터 및 신체 데이터가 추가적으로 더 이용됨으로써, 슈퍼에이저 여부와 관련되는 다양한 변수를 더 고려할 수 있다는 효과가 있다.

도면의 간단한 설명

[0029] 도 1은 본 개시의 일 실시예에 따른 슈퍼에이저 판단 장치의 블록도이다.

도 2는 본 개시의 일 실시예에 따른 슈퍼에이저 판단 장치를 이용하여 예측 인지검사 결과를 예측하는 방법을 나타낸 것이다.

도 3은 본 개시의 일 실시예에 따른 슈퍼에이저 판단 장치를 이용한 방법의 순서도이다.

도 4는 본 개시의 다른 실시예에 따른 슈퍼에이저 판단 장치의 블록도이다.

도 5는 본 개시의 다른 실시예에 따른 슈퍼에이저 판단 장치를 이용하여 인지검사 결과를 예측하는 방법을 나타낸 것이다.

도 6은 본 개시의 다른 실시예에 따른 슈퍼에이저 판단 장치를 이용한 방법의 순서도이다.

도 7은 본 개시의 일 실시예에 따른 주요균주를 선별하는 방법을 나타낸 것이다.

도 8은 본 개시의 일 실시예에 따른 학습모델의 종류에 따른 검증 결과를 나타낸 것이다.

도 9는 본 개시의 일 실시예에 따른 주요균주들의 SHAP value를 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0030] 이하, 본 개시의 일부 실시예들을 예시적인 도면을 통해 상세하게 설명한다. 각 도면의 구성 요소들에 참조 부호를 부가함에 있어서, 동일한 구성 요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 개시를 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 개시의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.

[0031] 본 개시에 따른 실시예의 구성요소를 설명하는 데 있어서, 제1, 제2, i), ii), a), b) 등의 부호를 사용할 수 있다. 이러한 부호는 그 구성요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 부호에 의해 해당 구성요소의 본질 또는 차례나 순서 등이 한정되지 않는다. 명세서에서 어떤 부분이 어떤 구성요소를 '포함' 또는 '구비'한다고 할 때, 이는 명시적으로 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0033] 본 개시에서, "피험자"는 마이크로바이옴 검사 대상이자, 인지능력 검사 대상이다.

[0034] 또한, 본 개시에서, "사용자"는 피험자의 슈퍼에이저 여부를 판단하고자 하는 자이며, 의료인, 간호인, 간병인 및 피험자 자신일 수 있다.

[0035] 또한, 본 개시에서, "마이크로바이옴 검사 데이터"는 피험자의 장내 마이크로바이옴을 구성하는 복수의 균주 및 각 균주들의 상대존재비(relative abundance)를 포함한다. 예를 들어, 마이크로바이옴검사 데이터는 Phylum, Class, Order, Family, Genus의 단위의 장내 균주들의 상대존재비를 포함할 수 있으나, 반드시 이에 한정되는 것은 아니며 다양한 단위의 상대존재비를 의미하는 것일 수 있다.

[0036] 또한, 본 개시에서, "인지검사 결과"는 피실험자의 인지기능 검사에 따른 결과를 의미하며, 바람직하게는 노년층의 인지기능을 검사하기 위한 검사 결과를 의미한다.

[0037] 또한, 본 개시에서, "예측 인지검사 결과"는 학습모델에 의해 예측되는 인지검사 결과를 의미한다.

[0038] 또한, 본 개시에서, "슈퍼에이저"는 피험자가 속한 그룹 내 평균 인지능력보다 더 우수한 인지능력을 갖는 피험자를 의미한다.

[0040] **슈퍼에이저 판단 장치 및 방법의 설명**

[0041] 일 실시예에 따른 슈퍼에이저 판단 장치의 설명

[0042] 도 1은 본 개시의 일 실시예에 따른 슈퍼에이저 판단 장치의 블록도이다.

[0043] 도 1을 참조하면, 본 개시의 일 실시예에 따른 슈퍼에이저 판단 장치(1)는, 학습정보 생성모듈(10) 및 학습모듈(20)을 포함한다.

[0044] 학습정보 생성모듈(10)은 입력받은 자료를 이용하여 학습정보를 생성하도록 구성된다. 이를 위하여, 학습정보 생성모듈(10)은 마이크로바이옴 검사 데이터 입력부(11), 인지검사 결과 입력부(12), 주요균주 연산모듈(13) 및 학습정보 데이터베이스(14)의 전부 또는 일부를 포함한다.

[0045] 마이크로바이옴 검사 데이터 입력부(11)는 마이크로바이옴 검사 데이터를 입력받도록 구성된다. 여기서, 마이크로바이옴 검사 데이터는 바람직하게는 피실험자의 장내 균주 및 장내 균주의 상대존재비를 포함한다. 한편, 마이크로바이옴 검사 데이터는 채변검사를 통해 취득될 수 있으나, 반드시 이에 한정되는 것은 아니다.

[0046] 인지검사 결과 입력부(12)는 인지검사 결과를 입력받도록 구성된다. 여기서, 인지검사 결과는 피실험자의 인지기능 검사 결과로서, 피실험자가 속한 연령 그룹의 평균 인지능력 이상의 인지능력을 갖는지 여부, 즉, 슈퍼에이저 해당 여부를 포함한다.

[0047] 주요균주 연산모듈(13)은 마이크로바이옴 검사 데이터 및 인지검사 결과를 이용하여 주요균주를 선별하도록 구성된다. 주요균주는 슈퍼에이저와 일반 고령자 집단 간 유의하게 차이가 있는 균주를 선별한 균주 군으로서, 주요균주를 선별하는 구체적인 방법에 관해서는, 도 7 내지 도 9에서 설명한다.

[0048] 학습정보 데이터베이스(14)는 주요균주의 상대존재비 및 인지검사 결과를 학습정보 데이터세트로서 저장한다.

[0049] 학습모듈(20)은 학습정보 데이터베이스(14)로부터 전송받은 학습정보 데이터세트를 이용하여, 학습모델을 학습시키도록 구성된다. 더 구체적으로, 학습모듈(20)은 주요균주의 상대존재비를 입력자료로 하고, 인지검사 결과를 출력자료로 하는 학습모델을 학습시킨다.

[0050] 여기서, 학습모델은, 바람직하게는 의사결정나무(Decision Tree)의 앙상블(Ensemble) 모델 중 배깅(Bagging) 방식의 RF(Random Forest)와 부스팅(Boosting) 방식인 GB(Gradient Boosting), CatBoost 중 어느 하나가 사용될 수 있다. 다만, 이에 한정되지 아니하며 SVM(Support Vector Machine), MLP(Multi Layer Perceptron) 등도 선택될 수 있다.

[0052] 도 2는 본 개시의 일 실시예에 따른 슈퍼에이저 판단 장치를 이용하여 예측 인지검사 결과를 예측하는 방법을 나타낸 것이다.

[0053] 도 1에 도시된 방법에 따라 학습모듈(20)은 학습모델을 학습시킨다. 도 2를 참조하면, 도 1에서 학습된 학습모델을 포함하는 학습모듈(20)에 사용자 단말기(미도시)를 통해 주요균주의 상대존재비가 입력되면, 상기한 주요균주의 상대존재비를 전송받아 학습시킨 학습모델에 입력하도록 구성된다. 학습모델로부터 예측 인지검사 결과가 출력되면, 학습모듈(20)은 출력되는 예측 인지검사 결과를 사용자 단말기(미도시)에 전송할 수 있다.

[0054] 예측 인지검사 결과에는 슈퍼에이저 해당 여부가 포함되어 있으며, 예측 인지검사 결과를 통해 피실험자가 슈퍼에이저에 해당하는지 확인될 수 있다.

[0055] 구체적으로, 사용자는 사용자 단말기(미도시)를 통해 수신한 예측 인지검사 결과를 통해 피험자의 슈퍼에이저 여부를 판단할 수 있다. 이때, 사용자가 직접 슈퍼에이저 여부를 판단할 수도 있고, 미리 마련된 장치 또는 모듈(미도시)을 통해 슈퍼에이저 여부가 판단될 수도 있으나, 어느 방식이든 무방하다.

[0056] 본 개시의 일 실시예에 따른 슈페에이저 판단 장치(1)에 따르면, 피험자의 마이크로바이옴 검사 데이터를 학습모듈(20)에 입력만 하여도, 피험자의 슈퍼에이저 해당 여부가 확인될 수 있어, 인지능력 평가에 대한 사용자들의 접근성이 높아진다는 효과가 있다.

[0057] 또한, 기존 인지검사 비용에 비해 월등히 낮은 비용으로 검사가 진행될 수 있어, 피험자의 경제적 부담이 완화된다는 효과가 있다.

[0058] 나아가, 기존 인지검사는 수 시간 이상의 검사 시간에 소요되는 데에 반해, 본 개시의 일 실시예에 따른 슈퍼에이저 판단 장치(1)를 이용하게 될 경우 피험자는 채변만 하면 되고, 사용자는 슈퍼에이저 판단 장치(1)를 구동하기만 하면 되므로, 그 과정이 간단하고, 시간 소요가 적다는 장점이 있다.

[0060] 일 실시예에 따른 슈퍼에이저 판단 방법의 설명

[0061] 도 3은 본 개시의 일 실시예에 따른 슈퍼에이저 판단 장치를 이용한 방법의 순서도이다.

[0062] 도 3을 참조하여, 본 개시의 일 실시예에 따른 슈퍼에이저 판단 장치(1)를 이용한 슈퍼에이저 판단 방법을 설명한다.

[0063] 마이크로바이옴 검사 데이터 입력부(11)를 통해 입력된 마이크로바이옴 검사 데이터와, 인지검사 결과 입력부(12)를 통해 입력된 인지검사 결과를 이용하여, 주요균주 연산모듈(13)이 주요균주를 선별한다(S300).

[0064] 이때, 마이크로바이옴 검사 데이터는, 바람직하게는 피험자의 채변 결과로부터 얻어질 수 있으며, 장내 균주의 종류 및 각 장내 균주의 상대존재비를 포함한다.

[0065] 인지검사 결과는, 바람직하게는 피험자의 인지능력 검사로부터 얻어질 수 있으며, 슈퍼에이저 여부를 포함하고 있다.

[0066] 주요균주 연산모듈(13)은 마이크로바이옴 검사 데이터 및 인지능력 검사를 이용하되, 미리 정해진 방법으로 연산하여 주요균주를 선별한다. 여기서, 미리 정해진 방법은 Mann-Whitney U test 및 Point Biserial Correlation 중 하나 이상을 포함할 수 있으나, 반드시 이에 한정되지 아니하며, 다른 상관도 측정 방법이 선택될 수 있다.

[0067] 주요균주 연산모듈(13)에 의해 선별되는 주요균주는 복수개이며, 바람직하게는 16개 이상일 수 있다. 주요균주와 관련하여, 도 7 내지 도 9에서 상세히 설명한다.

[0069] 주요균주 연산모듈(13)은 선별된 주요균주 각각의 상대존재비를 확인하고, 확인된 주요균주의 상대존재비와 인지검사 결과가 학습정보 데이터세트로서 학습정보 데이터베이스(14)에 저장된다(S310). 이때, 학습정보 데이터세트는 학습모델을 학습시키는 데에 필요한 데이터세트이다.

[0071] 학습모듈(20)은 학습정보 데이터베이스(14)에 정보를 요청하여, 학습정보 데이터세트를 전송받아 학습모델을 학습시킨다. 이때, 학습모델은 주요균주의 상대존재비를 입력자료로 하고, 인지검사 결과를 출력자료로 한다(S320).

[0072] 학습모델은 바람직하게는 의사결정나무(Decision Tree)의 앙상블(Ensemble) 모델 중 배깅(Bagging) 방식의 RF(Random Forest)와 부스팅(Boosting) 방식인 GB(Gradient Boosting), CatBoost 중 어느 하나가 사용될 수 있다. 다만, 이에 한정되지 아니하며 SVM(Support Vector Machine), MLP(Multi Layer Perceptron) 등도 선택될 수 있다.

[0074] 사용자 단말기(미도시)를 통해 입력된 주요균주의 상대존재비가 전달되어 학습모듈(20)에 입력되면, 학습모듈(20)은 예측 인지검사 결과를 출력할 수 있다(S330).

[0075] 이때, 예측 인지검사 결과에는 슈퍼에이저 해당 여부가 포함되어 있다. 예측 인지검사 결과는 사용자 단말기(미도시)에 예측 인지검사 결과를 전송될 수 있다.

[0077] S330 단계에서 출력되는 예측 인지검사 결과를 통해, 슈퍼에이저 여부가 확인될 수 있다(S340).

[0078] 사용자는 사용자 단말기(미도시)를 통해 수신한 예측 인지검사 결과를 통해, 피험자의 슈퍼에이저 여부를 판단

할 수 있다. 이때, 사용자가 직접 슈퍼에이저 여부를 판단할 수도 있고, 미리 마련된 장치 또는 모듈(미도시)을 통해 슈퍼에이저 여부가 판단될 수도 있으나, 어느 방식이든 무방하다.

[0079] 본 개시의 일 실시예에 따르면, 마이크로바이옴 검사 데이터를 기반으로 피험자의 인지능력, 더 구체적으로는 슈퍼에이저 여부가 평가됨으로써, 피험자가 인지능력 검사 시험지에 직접 답변을 체크한 경우보다 더 일관되며, 객관적인 결과가 제공된다는 장점이 있다.

[0080] 또한, 매번 다른 종류의 시험지가 제공될 필요가 없고, 마이크로바이옴 데이터만 취득되면 되기에, 시간에 따른 추적관찰이 가능한바, 지속적인 건강관리에 도움이 된다는 효과가 있다.

[0082] 다른 실시예에 따른 슈퍼에이저 판단 장치의 설명

[0083] 도 4는 본 개시의 다른 실시예에 따른 슈퍼에이저 판단 장치의 블록도이다.

[0084] 도 4를 참조하면, 본 개시의 다른 실시예에 따른 슈퍼에이저 판단 장치(1')는, 신상 데이터 입력부(15') 및 신체 데이터 입력부(16')를 더 포함한다.

[0085] 신상 데이터 입력부(15')는 신상 데이터를 입력받도록 구성된다. 여기서, 신상 데이터란, 피험자의 연령, 성별 및 교육 수준을 포함할 수 있으나, 반드시 이에 한정되는 것은 아니며, 소득 등을 더 포함할 수 있다.

[0086] 신체 데이터 입력부(16')는 신체 데이터를 입력받도록 구성된다. 여기서, 신체 데이터란, 체질량지수(BMI)를 포함할 수 있으나, 반드시 이에 한정되는 것은 아니며, 가족력 또는 병력 등을 더 포함할 수 있다.

[0087] 학습정보 생성모듈(10')은 마이크로바이옴 검사 데이터 입력부(11') 및 인지검사 결과 입력부(12') 각각에 입력된 정보들을 기반으로, 주요균주 연산모듈(13')에 의해 확인되는 주요균주의 상대존재비, 인지검사 결과, 신상 데이터 및 신체 데이터를 학습정보 데이터세트로 생성한다.

[0088] 생성된 학습정보 데이터세트는 학습정보 데이터베이스(14')에 저장된다.

[0089] 학습정보 데이터베이스(14')에 저장된 학습정보 데이터세트는 학습모듈(20')에 전송될 수 있다.

[0091] 도 5는 본 개시의 다른 실시예에 따른 슈퍼에이저 판단 장치를 이용하여 인지검사 결과를 예측하는 방법을 나타낸 것이다.

[0092] 도 4에 도시된 방법에 따라 학습된 학습모델을 포함하는 학습모듈(20')에, 사용자 단말기(미도시)를 통해 입력된 주요균주의 상대존재비, 신상 데이터 및 신체 데이터가 입력되면, 예측 인지검사 결과가 출력된다.

[0093] 출력되는 예측 인지검사 결과는, 사용자 단말기(미도시)에 전송될 수 있고, 사용자는 예측 인지검사 결과를 이용하여 피험자의 슈퍼에이저 여부를 확인할 수 있다.

[0095] 도 6은 본 개시의 다른 실시예에 따른 슈퍼에이저 판단 장치를 이용한 방법의 순서도이다.

[0096] 도 6을 참조하여, 본 개시의 다른 실시예에 따른 슈퍼에이저 판단 장치(1')를 이용한 슈퍼에이저 판단 방법을 설명한다.

[0097] 마이크로바이옴 검사 데이터 입력부(11')를 통해 입력된 마이크로바이옴 검사 데이터와, 인지검사 결과 입력부(12')를 통해 입력된 인지검사 결과를 이용하여, 주요균주 연산모듈(13')이 주요균주를 선별한다(S600).

[0099] 주요균주 연산모듈(13')은 선별된 주요균주 각각의 상대존재비를 확인한다. 이후에, 확인된 주요균주의 상대존재비, 인지검사 결과, 신상 데이터 입력부(15')를 통해 입력된 신상 데이터 및 신체 데이터 입력부(16')를 통해 입력된 신체 데이터가 학습정보 데이터세트로서 학습정보 데이터베이스(14')에 저장된다(S610).

[0101] 학습모듈(20')은 학습정보 데이터베이스(14')에 정보를 요청하여, 학습정보 데이터세트를 전송받아 학습모델을 학습시킨다. 이때, 학습모델은 주요균주의 상대존재비, 신상 데이터, 신체 데이터를 입력자료로 하고, 인지검사

결과를 출력자료로 한다(S620).

[0103] 사용자 단말기(미도시)를 통해 주요균주의 상대존재비, 신상 데이터 및 신체 데이터가 전달되어 학습모듈(20')에 입력되면, 학습모듈(20')은 예측 인지검사 결과를 출력할 수 있다(S630).

[0105] S630 단계에서 출력되는 예측 인지검사 결과를 통해, 슈퍼에이저 여부가 확인될 수 있다(S640).

[0106] 본 개시의 다른 실시예에 따른 슈퍼에이저 판단 방법은, 신상 데이터 및 신체 데이터를 추가적으로 이용함으로써, 보다 다양한 변수를 고려할 수 있다는 장점이 있다.

[0108] 이상에서는, 본 개시의 실시예들에 따른 슈퍼에이저 판단 장치 및 방법의 일반론을 설명하였으며, 이하에서는 본 개시의 일 실시예에 따른 구체적인 균주 선별 예시 및 학습모델 선정 등과 관련하여 설명한다.

[0110] **학습정보 생성방법 및 검증 결과**

[0111] 피실험자의 구성 및 데이터 분류

[0112] 피실험자 그룹은 만 60세 이상의 고령자 102명으로 구성되었으며, 모든 피실험자에 대하여 마이크로바이옴 검사 및 인지기능 검사가 수행되었다.

[0113] 피실험자 그룹은 인지기능 검사 결과, 본인이 속한 연령그룹의 평균 인지능력 이상의 인지능력을 가지는 57명의 슈퍼에이저 집단과 45명의 일방 고령자 집단으로 분류되었다.

[0114] 이때, 학습모델에 사용된 데이터는 71개이며, 검증데이터로 31개가 사용되었으며, hold-out 방식으로 학습 및 성능 검증이 수행되었다.

[0116] 주요균주 선별방법

[0117] 본 개시의 일 실시예에 따르면, 주요균주 연산모듈(13)은 주요균주를 선별할 때, Mann-Whitney U test(도 7의 (a) 참조) 및 Point Biserial Correlation(도 7의 (b) 참조) 방법을 이용하여 주요균주를 선별할 수 있다.

[0118] 도 7은 본 개시의 일 실시예에 따른 주요균주를 선별하는 방법을 나타낸 것이다. 구체적으로, 도 7의 (a)는 Mann-Whitney U test 방법을 통해 유의한 것으로 확인된 균주를 나타낸 것이다.

[0119] Mann-Whitney U test 방법은 슈퍼에이저 집단과 일반 고령자 집단 간 유의하게 차이나는 균주를 선별하기 위한 방법이다.

[0120] 도 7의 (a)를 참조하면, Phylum 'Firmicutes'와 'Actinobacteria'에서 유의한 균주가 선별되었다.

[0121] 또한, Firmicutes의 Class 'Clostridia'는 AC001115_g, Eisenbergiella, PAC001138_g, Sellimonas, PAC001236_g, Romboutsia, Eubacterium_g8, Subdoligranulum 8개의 균주가 타균주에 비하여 유의하게 슈퍼에이저 여부와 상관관계가 있는 것으로 확인되었으며, Class 'Bacilli'에서는 Family 'Lactobacillaceae'와 Genus 'Leuconostoc'이, Phylum 'Bacteroidetes'에서는 Family 'Odoribacteraceae'이 타균주에 비하여 유의하게 슈퍼에이저 여부와 상관관계가 있는 것으로 확인되었다.

[0122] 이때, Sellimonas와 Romboutsia를 제외한 9개의 균주가 일반 고령자 집단보다 슈퍼에이저 집단에서 유의하게 더 큰 값을 보였다.

[0123] 정리하면, Mann-Whitney U test 방법을 통해 11개의 주요균주가 선별되었다.

[0125] 도 7의 (b)는 Point Biserial Correlation 방법을 통해 유의한 것으로 확인된 균주를 나타낸 것이다.

[0126] Point Biserial Correlation 방법에 따르면, 슈퍼에이저와 일반 고령자의 라벨을 각각 1과 0의 범주형 변수로

두고, 슈퍼에이저 여부(즉, 1)와 유의하게 상관관계가 있는 주요균주가 추출된다. 여기서, 유의한 상관관계란 p value가 0.05 미만인 것을 의미한다.

[0127] 도 7의 (b)를 참조하면, Phylum 'Firmicutes'와 'Actinobacteria'에서 유의한 균주가 확인되었다.

[0128] 또한, Firmicutes의 Clostridia의 Genus Coprococcus, Ruminococcus_g4, Sellimonas, PAC001236_g, Eubacterium_g8과 Actinobacteria와 이의 Class Actinobacteria_c, 그리고 Genus인 Rothia가 $p < 0.05$ 범위에 해당하여, 유의한 균주로 선별되었다.

[0129] 또한, Firmicutes의 Family인 Lachnospiraceae의 균주인 Coprococcus, Ruminococcus_g4, Sellimonas는 음의 상관관계를 보이며, 이를 제외한 주요균주들은 양의 상관관계를 보이는 것을 확인하였다.

[0130] 정리하면, Point Biserial Correlation 방법을 통해 8개의 주요균주가 선별되었으며, 그 중 3개는 Mann-Whitney U test 방법을 통해 선별된 주요균주와 중복된다. 따라서, 본 개시에서는 총 16개의 주요균주가 선별되었다.

[0131] 한편, 주요균주 선별 시 사용되는 선별 방법의 종류에 따라, 주요균주의 종류 및 개수가 상이해질 수 있음에 유의한다.

[0133] 학습모델 선정방법

[0134] 도 8은 본 개시의 일 실시예에 따른 학습모델의 종류에 따른 검증 결과를 나타낸 것이다.

[0135] <표 1>은 학습모델 별 성능(performance)을 나타낸 것이다.

표 1

[0137]

Methods	Accuracy	Precision	Recall	AUC	F1 Score
RF	0.871	0.842	0.941	0.863	0.889
GB	0.806	0.824	0.824	0.805	0.824
CatBoost	0.806	0.867	0.765	0.811	0.812
RF, Random Forest; GB, Gradient Boosting.					

[0139] 도 8 및 <표 1>을 참조하면, 의사결정나무 기반 앙상블인 RF 모델의 경우 Accuracy 87.1%, 슈퍼에이저에 대한 Recall 94.1%, AUC 86.3%로 높은 성능을 보였으며, 동일한 데이터로 GB를 학습하고 검증한 결과 Accuracy, Precision, Recall, AUC 등에서 80% 이상의 성능을 보였다.

[0140] 이에 따라, 본 개시에서는 의사결정나무 기반 앙상블인 RF 모델이 학습모델로서 사용되는 것이 바람직하다.

[0142] 검증 결과

[0143] 도 9는 본 개시의 일 실시예에 따른 주요균주들의 SHAP value를 나타낸 것이다. 구체적으로, 도 9에 따른 결과는 도 7에서 확인된 16개의 주요균주 및 도 8에서 확인된 RF 모델을 이용하여 학습모듈(20)이 학습모델을 학습한 결과로서, 이를 통해 균주들의 중요도가 확인될 수 있다.

[0144] 도 9a는 Test data에 RF 모델을 적용시켰을 때 각 균주들의 SHAP value를 나타내며, 도 9b는 각 균주가 예측값에 어떻게 영향을 미쳤는지를 나타낸다.

[0145] 예를 들어, Actinobacteria의 경우 값이 빨간색인 경우 SHAP value가 양의 값을 갖고 파란색인 경우 음의 값을 가진다. 이는 이 균주의 값이 높아질수록 1인 슈퍼에이저로 분류되는데 긍정적인 영향을 미치고, 낮을수록 부정적인 영향, 즉 일반 고령자(typical-ager)로 분류되도록 영향을 미친다는 것이다.

[0146] 도 9a 내지 도 9b를 참조하면, Actinobacteria, Eubacterium_g8, Subdoligranulum, Eisenbergiella, PAC001115_g, PAC001138_g, Leuconostoc은 값이 높을수록 슈퍼에이저로 분류되도록 작용하는 경향이 있으며,

Romboutsia, Ruminococcus_g4 이와 반대로 작용하는 경향이 있음을 알 수 있다.

[0147] 도 9c는 도 9a 및 도 9b의 결과를 정리한 것이다.

[0148] 도 9c의 'Statistics Analysis Result'열에서는, U-test와 Correlation에서 슈퍼에이저의 집단이 유의하게 높은 값을 갖거나 균주와 슈퍼에이저 여부가 유의한 양의 관계를 갖는 경우 빨간색으로 표현되었고, 유의하게 낮거나 유의하게 음의 관계를 갖는 경우 파란색으로 표현되었다.

[0149] 또한, 'SHAP value'열에서는, 슈퍼에이저로 분류되는데 긍정적인 영향을 미친 경우 빨간색 +, 부정적인 영향을 미친 경우 파란색 -로 표현되었다.

[0150] 마지막 'Mean(|SHAP value|)' 열에서는, SHAP value의 절대값의 평균으로 학습모델에서 균주의 중요도(영향도)가 확인될 수 있다.

[0151] 도 9c를 참조하면, SHAP value가 슈퍼이에저 여부에 영향을 미치는 방향 및 선별방식(Mann-Whitney U test 및 Point Biserial Correlation)의 방향이 일치함을 알 수 있다.

[0152] 이상의 설명은 본 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 실시예들은 본 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

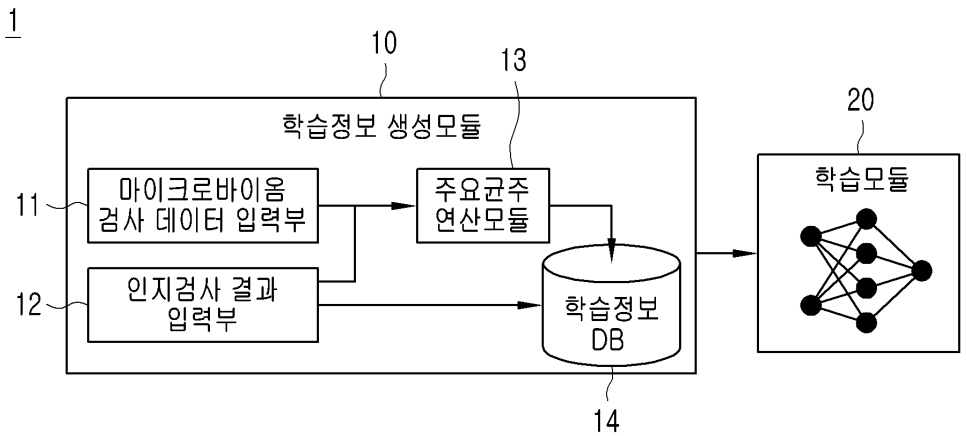
[0154] 1, 1': 슈퍼이에저 판단 장치

10, 10': 학습정보 생성모듈

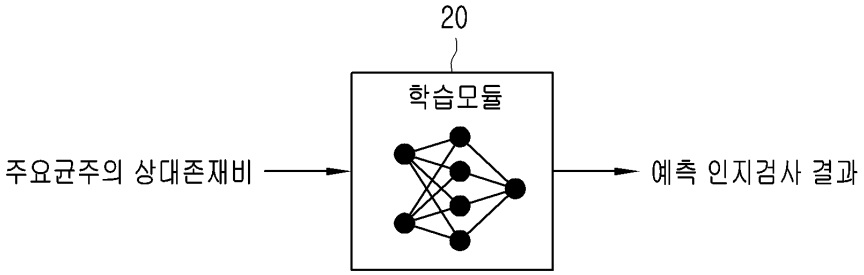
20, 20': 학습모듈

도면

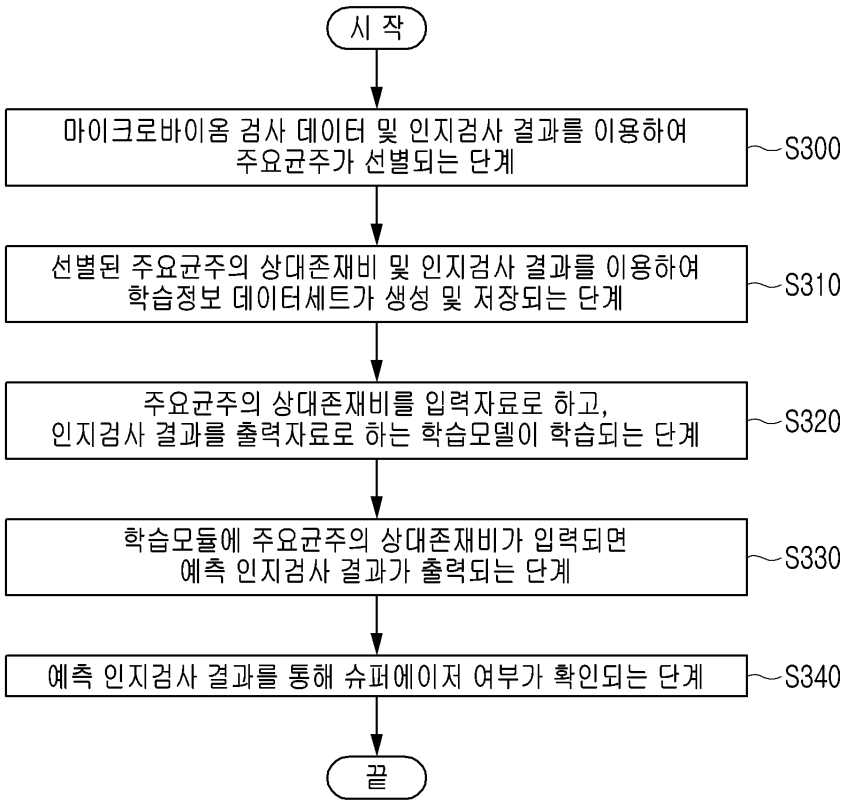
도면1



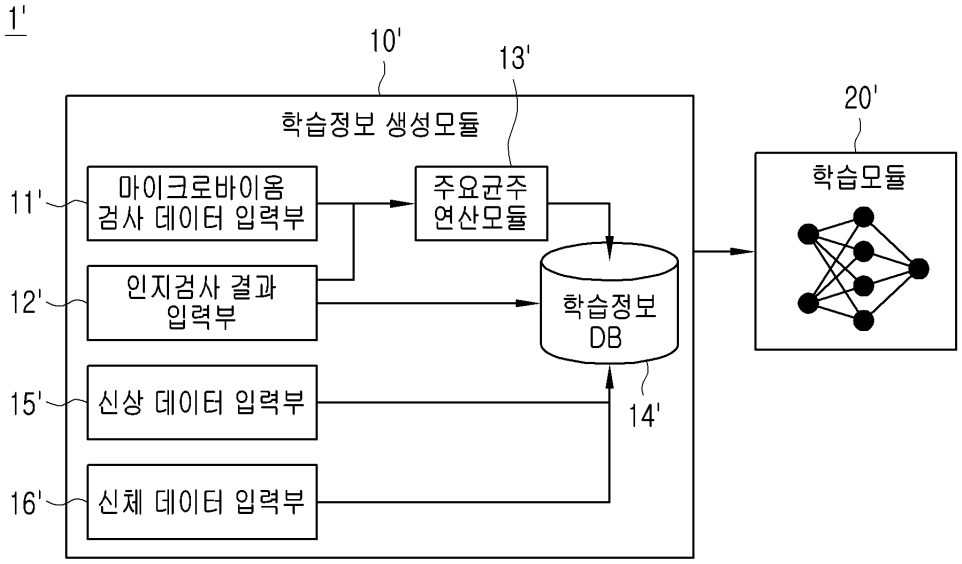
도면2


20
학습모듈
주요균주의 상대존재비
예측 인지검사 결과

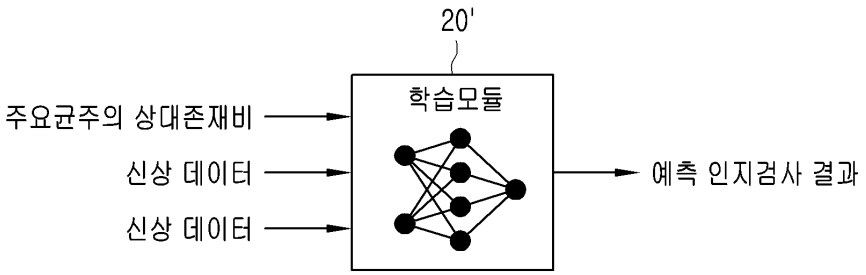
도면3


시 작
마이크로바이옴 검사 데이터 및 인지검사 결과를 이용하여 주요균주가 선별되는 단계
S300
선별된 주요균주의 상대존재비 및 인지검사 결과를 이용하여 학습정보 데이터세트가 생성 및 저장되는 단계
S310
주요균주의 상대존재비를 입력자료로 하고, 인지검사 결과를 출력자료로 하는 학습모델이 학습되는 단계
S320
학습모듈에 주요균주의 상대존재비가 입력되면 예측 인지검사 결과가 출력되는 단계
S330
예측 인지검사 결과를 통해 슈퍼에이저 여부가 확인되는 단계
S340
끝

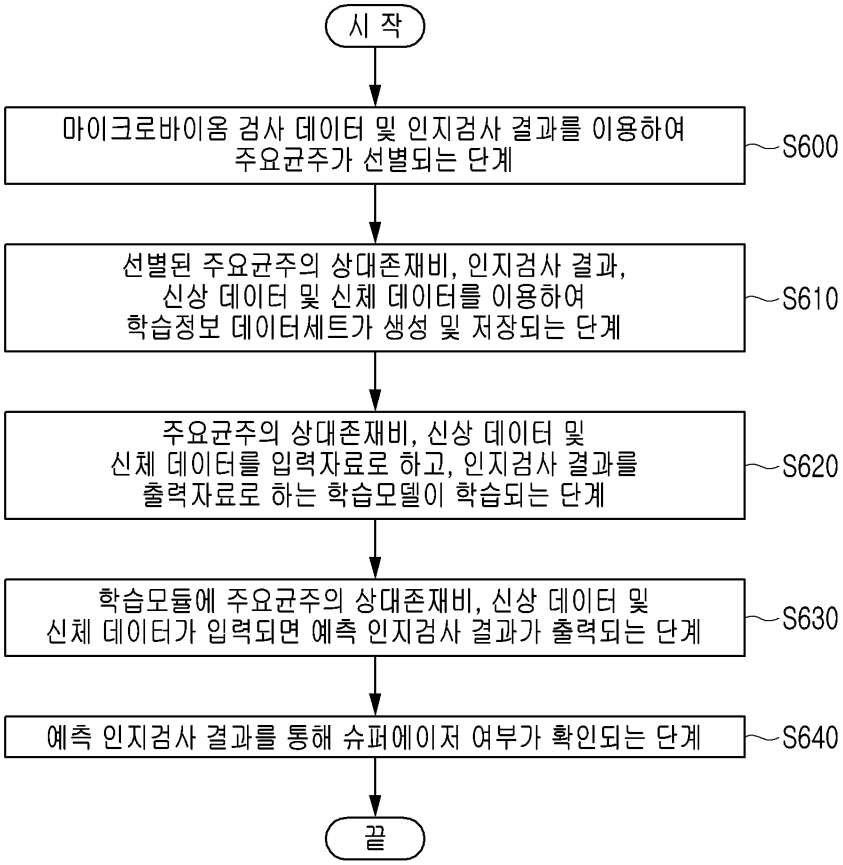
도면4


1'
10'
13'
학습정보 생성모듈
20'
학습모듈
11'
마이크로바이옴
검사 데이터 입력부
주요균주
연산모듈
12'
인지검사 결과
입력부
학습정보
DB
15'
신상 데이터 입력부
14'
16'
신체 데이터 입력부

도면5


20'
학습모듈
주요균주의 상대존재비
신상 데이터
신상 데이터
예측 인지검사 결과

도면6


시 작
마이크로바이옴 검사 데이터 및 인지검사 결과를 이용하여 주요균주가 선별되는 단계
S600
선별된 주요균주의 상대존재비, 인지검사 결과, 신상 데이터 및 신체 데이터를 이용하여 학습정보 데이터세트가 생성 및 저장되는 단계
S610
주요균주의 상대존재비, 신상 데이터 및 신체 데이터를 입력자료로 하고, 인지검사 결과를 출력자료로 하는 학습모델이 학습되는 단계
S620
학습모듈에 주요균주의 상대존재비, 신상 데이터 및 신체 데이터가 입력되면 예측 인지검사 결과가 출력되는 단계
S630
예측 인지검사 결과를 통해 슈퍼에이저 여부가 확인되는 단계
S640
끝

도면7

Phylum	Class	Order	Family	Genus	Mean		U-test	
					SuperAger	TypicalAger	Statistic	p-value
Firmicutes	Clostridia	Clostridiales	Christensenellaceae	**PAC001115_g**	1.2426	0.7663	1568.00	0.05
			Lachnospiraceae	**Eisenbergiella**	1.2426	0.7663	1617.00	0.02
				PAC001138_g	0.0422	0.0345	1628.00	0.02
				Sellimonas	0.0303	0.2628	980.00	0.04
			Mogibacterium_f	**PAC001236_g**	0.0570	0.0260	1628.00	0.02
			Peptostreptococcaceae	**Romboutsia**	0.3693	0.7411	988.00	0.05
			Ruminococcaceae	**Eubacterium_g8**	1.2263	0.5073	1703.50	0.00
				Subdoligranulum	0.7770	0.5720	1623.50	0.02
	Bacilli	Lactobacillales	**Lactobacillaceae**		2.7559	1.0275	1640.00	0.02
			Leuconostocaceae	**Leuconostoc**	2.7853	1.0136	1698.50	0.01
Bacteroidetes	Bacteroidia	Bacteroidales	**Odoribacteraceae**		0.1583	0.0887	1595.00	0.03

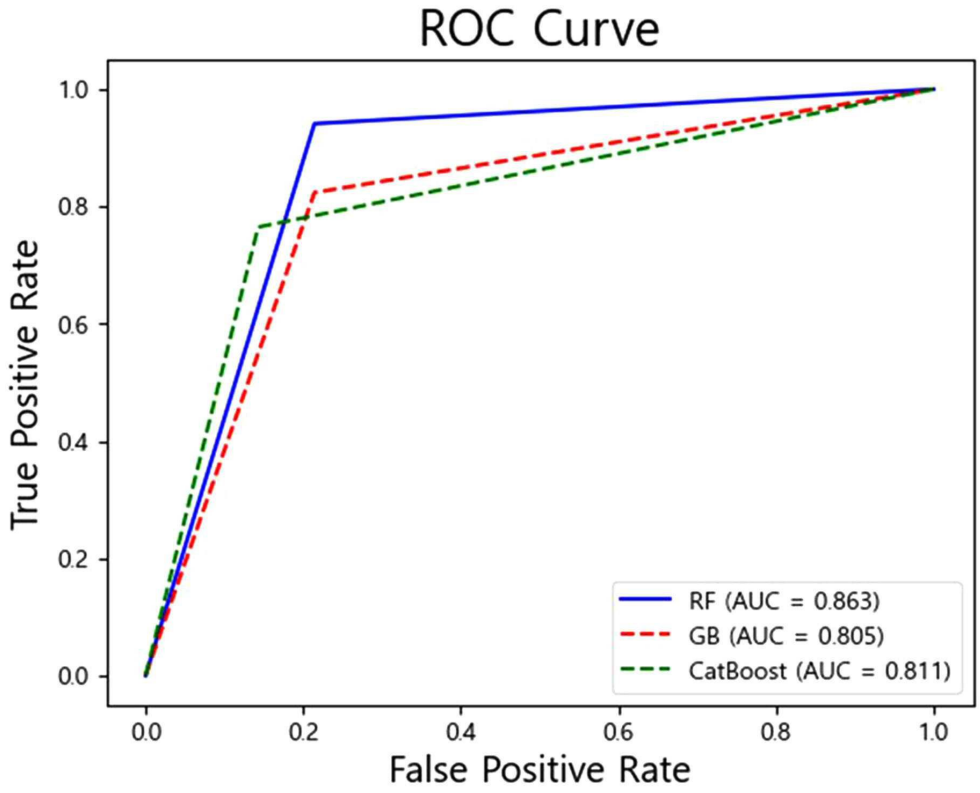
(a)

*U-test에서 유의하게 나온 균주

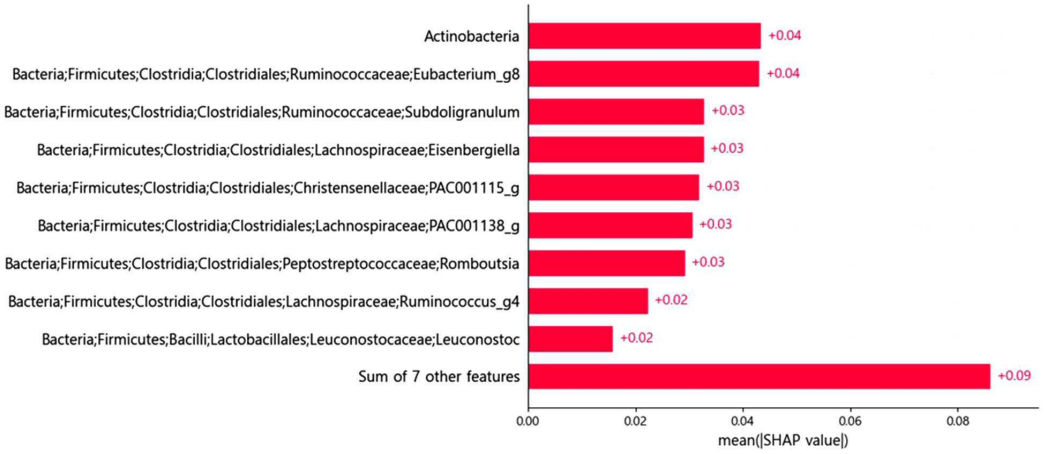
Phylum	Class	Order	Family	Genus	Correlation	
					Coefficient	P-value
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	**Coprococcus**	-0.228	0.02
				Ruminococcus_g4	-0.244	0.01
				Sellimonas	**-0.272**	0.01
			Mogibacterium_f	**PAC001236_g**	0.205	0.04
			Ruminococcaceae	**Eubacterium_g8**	**0.284**	0.00
Actinobacteria					0.266	0.01
	Actinobacteria_c				0.199	0.04
		Micrococcales	Micrococcaceae	**Rothia**	0.217	0.03

(b)

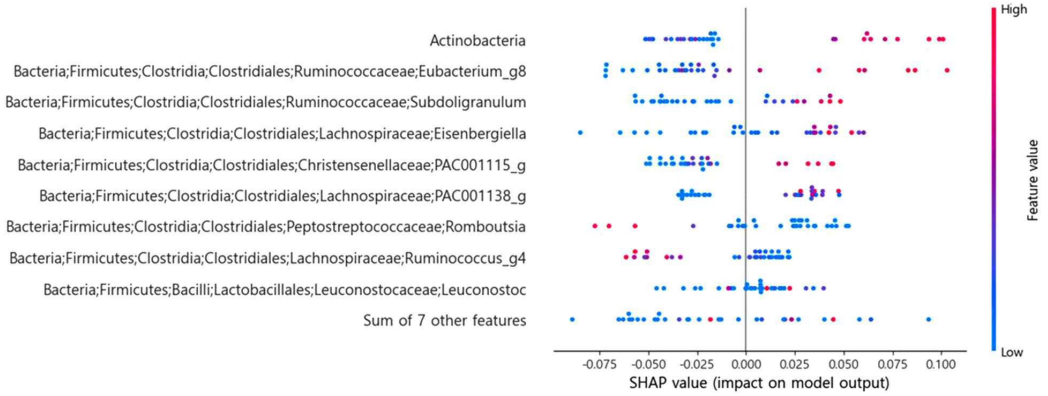
도면8


ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
RF (AUC = 0.863)
GB (AUC = 0.805)
CatBoost (AUC = 0.811)

도면9a


Actinobacteria
Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Eubacterium_g8
Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Subdoligranulum
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Eisenbergiella
Bacteria;Firmicutes;Clostridia;Clostridiales;Christensenellaceae;PAC001115_g
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;PAC001138_g
Bacteria;Firmicutes;Clostridia;Clostridiales;Peptostreptococcaceae;Romboutsia
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Ruminococcus_g4
Bacteria;Firmicutes;Bacilli;Lactobacillales;Leuconostocaceae;Leuconostoc
Sum of 7 other features
+0.04
+0.04
+0.03
+0.03
+0.03
+0.03
+0.03
+0.02
+0.02
+0.09
0.00
0.02
0.04
0.06
0.08
mean(|SHAP value|)

도면9b



도면9c

Feature					Statistics Analysis Result	Shap Value (impact 방향)	Mean(|SHAP value|)
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Eubacterium_g8	[U-test] significantly higher [Correlation] Positive correlation(0.284)	+	0.04
				Subdoligranulum	[U-test] significantly higher	+	0.03
			Lachnospiraceae	Eisenbergiella	[U-test] significantly higher	+	0.03
				Ruminococcus_g4	[Correlation] Negative correlation(-0.244)	-	0.02
				Sellimonas	[U-test] significantly lower [Correlation] Negative correlation(-0.272)		
				PAC001138_g	[U-test] significantly higher	+	0.03
				Coprococcus	[Correlation] Negative correlation(-0.228)		
			Christensenellaceae	PAC001115_g	[U-test] significantly higher [Correlation] Positive correlation(0.205)	+	0.03
			Peptostreptococcaceae	Romboutsia	[U-test] significantly lower	-	0.03
			Mogibacterium_f	PAC001236_g	[U-test] significantly higher		
	Bacilli	Lactobacillales	Lactobacillaceae		[U-test] significantly higher		
			Leuconostocaceae	Leuconostoc	[U-test] significantly higher	+	0.02
Actinobacteria					[Correlation] Positive correlation(0.266)	+	0.04
	Actinobacteria_c				[Correlation] Positive correlation(0.199)		
		Micrococcales	Micrococcaceae	Rothia	[Correlation] Positive correlation(0.217)		
Bacteroidetes	Bacteroidia	Bacteroidales	Odoribacteraceae		[U-test] significantly higher		