(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0158948
(43) 공개일자 2022년12월02일

(51) 국제특허분류(Int. Cl.)
G16H 50/70 (2018.01) *G06N 20/00* (2019.01)
G16B 40/20 (2019.01) *G16H 50/20* (2018.01)
G16H 50/30 (2018.01) *G16H 50/50* (2018.01)
(52) CPC특허분류
G16H 50/70 (2018.01)
G06N 20/00 (2021.08)
(21) 출원번호 10-2021-0066612
(22) 출원일자 2021년05월25일
심사청구일자 2021년05월25일

(71) 출원인
주식회사 에이치이엠파마
경상북도 포항시 북구 흥해읍 한동로 558, 204호, 에프101호, 에프102호, 102호, 205호, 306호(창업보육센터)
(72) 발명자
지요셉
경기도 수원시 영통구 웰빙타운로 70, 8707동 1503호(이의동, 광교 호반베르디움)
박소영
경기도 수원시 영통구 광교중앙로 145, A동 1018호(이의동)
(74) 대리인
특허법인엠에이피에스

전체 청구항 수 : 총 14 항

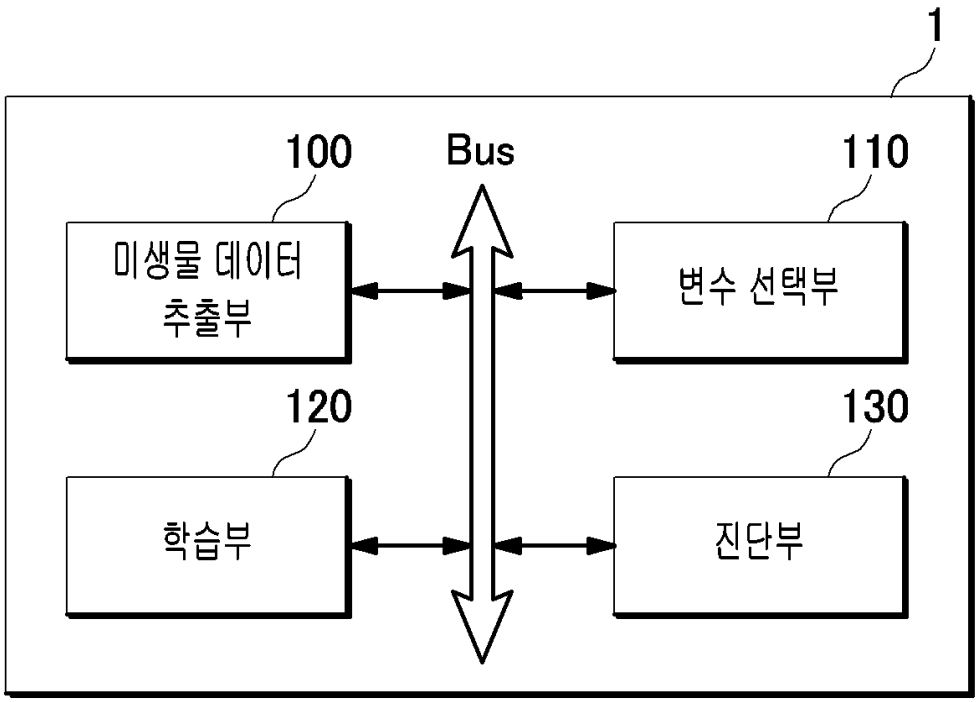
(54) 발명의 명칭 **머신러닝 모델을 이용하여 복부팽만 유무를 판별하는 방법 및 진단 장치**

(57) 요 약

머신러닝 모델을 이용하여 복부팽만 유무를 판별하는 방법은 개체로부터 채취한 장내 유래 물질을 장내 환경 유사 조성물과 혼합한 혼합물을 분석하는 단계, 상기 혼합물의 분석 결과에 기초하여 복수의 미생물 데이터를 추출하는 단계, 기설정된 변수 선택 알고리즘에 기초하여 상기 복수의 미생물 데이터 중 머신러닝 모델에 사용될 미

(뒷면에 계속)

대 표 도 - 도1



생물 관련 변수를 선택하는 단계, 상기 미생물 관련 변수를 이용하여 상기 머신러닝 모델을 학습시키는 단계 및 검사 대상 객체로부터 채취한 미생물 데이터를 상기 학습된 머신러닝 모델에 입력하여 복부팽만 유무를 판별하는 단계를 포함할 수 있다. 상기 미생물 관련 변수는 클로스트리디아세에 (Clostridiaceae), 리케넬라세에 (Rikenellaceae), 덴테로코카세에 (Enterococcaceae), 박테로이다세에 (Bacteroidaceae), 데리시펠로트리카세에 (Erysipelotrichaceae), 류코노스토카세에 (Leuconostocaceae), 마리니필라세에 (Marinifilaceae), 키리스텐세넬라세에 (Christensenellaceae), 타네렐라세에 (Tannerellaceae), 락토바실라세에 (Lactobacillaceae), 펩토스트렙토코카세에 (Peptostreptococcaceae), 비피도박테리아세에 (Bifidobacteriaceae), 데플루비탈레세에 (Defluviitaleaceae), 코리오박테리아세에 (Coriobacteriaceae), 마이크로코카세에 (Micrococcaceae), 루미노코카세에 (Ruminococcaceae), 오실로스피라세에 (Oscillospiraceae), 라크노스피라세에 (Lachnospiraceae), 수테렐라세에 (Sutterellaceae), 엑티노미세타세에 (Actinomycetaceae), 에게텔라세에 (Eggerthellaceae), 펩토스트렙토코칼레스-티시에렐랄레스 (Peptostreptococcales-Tissierellales), 아케르만시아세에 (Akkermansiaceae) 과(Family)에 속하는 속(Genus)에서 선택되는 1종 이상의 미생물의 함량을 포함할 수 있다.

(52) CPC특허분류
G16B 40/20 (2019.02)
G16H 50/20 (2018.01)
G16H 50/30 (2018.01)
G16H 50/50 (2018.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1415170308
과제번호 P0009477
부처명 산업통상자원부
과제관리(전문)기관명 한국산업기술진흥원
연구사업명 사업화연계기술개발사업
연구과제명 머신러닝을 이용하여 기능성이 개선된 개인 맞춤형 면역 개선 개별인정형 프로바이오틱스 개발
기 여 율 1/1
과제수행기관명 주식회사 에이치이엠
연구기간 2019.04.01 ~ 2021.03.31

명 세 서

청구범위

청구항 1

머신러닝 모델을 이용하여 복부팽만 유무를 판별하는 방법에 있어서,

개체로부터 채취한 장내 유래 물질을 장내 환경 유사 조성물과 혼합한 혼합물을 분석하는 단계;

상기 혼합물의 분석 결과에 기초하여 복수의 미생물 데이터를 추출하는 단계;

기설정된 변수 선택 알고리즘에 기초하여 상기 복수의 미생물 데이터 중 머신러닝 모델에 사용될 미생물 관련 변수를 선택하는 단계;

상기 미생물 관련 변수를 이용하여 상기 머신러닝 모델을 학습시키는 단계; 및

검사 대상 객체로부터 채취한 미생물 데이터를 상기 학습된 머신러닝 모델에 입력하여 복부팽만 유무를 판별하는 단계

를 포함하고,

상기 미생물 관련 변수는 클로스트리디아세에 (Clostridiaceae), 리케넬라세에 (Rikenellaceae), 덴테로코카세에 (Enterococcaceae), 박테로이다세에 (Bacteroidaceae), 데리시펠로트리카세에 (Erysipelotrichaceae), 류코노스토카세에 (Leuconostocaceae), 마리니필라세에 (Marinifilaceae), 키리스텐세넬라세에 (Christensenellaceae), 타네렐라세에 (Tannerellaceae), 락토바실라세에 (Lactobacillaceae), 펩토스트렙토코카세에 (Peptostreptococcaceae), 비피도박테리아세에 (Bifidobacteriaceae), 데플루비탈레세에 (Defluviitaleaceae), 코리오박테리아세에 (Coriobacteriaceae), 마이크로코카세에 (Micrococcaceae), 루미노코카세에 (Ruminococcaceae), 오실로스피라세에 (Oscillospiraceae), 라크노스피라세에 (Lachnospiraceae), 수테렐라세에 (Sutterellaceae), 엑티노미세타세에 (Actinomycetaceae), 에게텔라세에 (Eggerthellaceae), 펩토스트렙토코칼레스-티시에렐랄레스 (Peptostreptococcales-Tissierellales), 아케르만시아세에 (Akkermansiaceae) 과(Family)에 속하는 속(Genus)에서 선택되는 1종 이상의 미생물의 함량을 포함하는 것인, 복부팽만 유무 판별 방법.

청구항 2

제 1 항에 있어서,

상기 혼합물을 분석하는 단계는,

상기 혼합물을 혐기 조건에서 18시간 내지 24시간 동안 배양하는 단계; 및

상기 혼합물이 배양된 배양물을 분석하는 단계

를 포함하는 것인, 복부팽만 유무 판별 방법.

청구항 3

제 2 항에 있어서,

상기 배양물을 분석하는 단계는,

상기 배양물을 원심 분리하여 상등액과 침전물을 분리한 후, 상기 상등액 및 상기 침전물을 분석하는 단계

를 포함하는 것인, 복부팽만 유무 판별 방법.

청구항 4

제 2 항에 있어서,

상기 미생물 데이터는 상기 배양물에 포함된 내독소(endotoxin), 황화수소(hydrogen sulfide), 단쇄지방산(Short-chain fatty acids, SCFAs) 및 장균총 유래 대사체 중 하나 이상의 함량, 농도, 종류, 장균총에 포함된 균의 종류, 농도, 함량 또는 다양성 변화 중 적어도 하나를 포함하는 것인, 복부팽만 유무 판별 방법.

청구항 5

제 1 항에 있어서,

상기 변수 선택 알고리즘은 보루타(Boruta) 알고리즘, 재귀 변수 제거(RFE: Recursive Feature Elimination) 알고리즘 중 적어도 하나를 포함하는 것인, 복부팽만 유무 판별 방법.

청구항 6

제 1 항에 있어서,

상기 머신러닝 모델은 로지스틱회귀(Logistic Regression) 모델, GLMNET(Generalized linear) 모델, 랜덤포레스트(Random Forest) 모델, 그래디언트 부스팅(Gradient Boosting) 모델, XGB(Extreme Gradient Boosting) 모델 중 적어도 하나를 포함하는 것인, 복부팽만 유무 판별 방법.

청구항 7

제 1 항에 있어서,

상기 미생물 관련 변수는 클로스트리디엄 (Clostridium), 서브도리그라눌럼 (Subdoligranulum), 알리스티페스 (Alistipes), 플라보니프랙터 (Flavonifractor), 블라우티아 (Blautia), 우에이셀라 (Weissella), 수테렐라 (Sutterella), 파라박테로이데스 (Parabacteroides), 락토바실러스 (Lactobacillus), 롬보웃시아 (Romboutsia), 인테스티니박터 (Intestinibacter), 루미노코커스 (Ruminococcus), 비피도박테리엄 (Bifidobacterium), 안에어로스티페스 (Anaerostipes), 도레아 (Dorea), 데플루비이탈레세에 (Defluviitaleaceae_UCG-011), 엑티노마이세스 (Actinomyces), 유박테리엄 (Eubacterium), 에게텔라 (Eggerthella), 류코노스톡 (Leuconostoc), 라크노스피라세에 (Lachnospiraceae_FCS020_group), 라크노스피라 (Lachnospira), 안에어로코커스 (Anaerococcus), 펩토니필러스 (Peptoniphilus), 테리스포로박터 (Terrisporobacter), 셀리모나스 (Sellimonas), 슬라키아 (Slackia), UCG-003, 캔디다터스_솔레아페리아 (Candidatus_Soleaferrea), 안에어로스티그넘 (Anaerostignum), 박테로이데스 (Bacteroides), UCG-002, 엔테로코커스 (Enterococcus), 투리시박터 (Turicibacter), 푸시카테니박터 (Fusicatenibacter), 아케르만시아 (Akkermansia), 튜리시박터 (Turicibacter) 속(Genus)에 속하는 종(Species)에서 선택되는 1종 이상의 미생물의 함량을 포함하는 것인, 복부팽만 유무 판별 방법.

청구항 8

제 1 항에 있어서,

상기 미생물 관련 변수는 페르프링겐스 (Perfringens), 스하히이 (Shahii), 디스타소니스 (Distasonis), 포르미시게네란스 (Formicigenerans), 테타이오타오미크론 (Thetaiotaomicron), 피네골디이 (Finegoldii), 메르데 (Merdae), 온데르동키이 (Onderdonkii), 부티리쿰 (Butyricum), 롱검 (Longum), 문드티이 (Mundtii), 파라푸트리피쿰 (Paraputrificum), 락티스 (Lactis), 애니말리스 (Animalis), 무시니필라 (Muciniphila) 종(Species)에서 선택되는 1종 이상의 미생물의 함량을 포함하는 것인, 복부팽만 유무 판별 방법.

청구항 9

머신러닝 모델을 이용하여 복부팽만 유무를 진단하는 장치에 있어서,

개체로부터 채취한 장내 유래 물질을 장내 환경 유사 조성물과 혼합한 혼합물의 분석 결과에 기초하여 복수의 미생물 데이터를 추출하는 미생물 데이터 추출부;

기설정된 변수 선택 알고리즘에 기초하여 상기 복수의 미생물 데이터 중 머신러닝 모델에 사용될 미생물 관련 변수를 선택하는 변수 선택부;

상기 미생물 관련 변수를 이용하여 상기 머신러닝 모델을 학습시키는 학습부; 및

검사 대상 객체로부터 채취한 미생물 데이터를 상기 학습된 머신러닝 모델에 입력하여 복부팽만 유무를 진단하는 진단부

를 포함하고,

상기 미생물 관련 변수는 클로스트리디아세에 (Clostridiaceae), 리케넬라세에 (Rikenellaceae), 덴테로코카세에 (Enterococcaceae), 박테로이다세에 (Bacteroidaceae), 데리시펠로트리카세에 (Erysipelotrichaceae), 류코노스토카세에 (Leuconostocaceae), 마리니필라세에 (Marinifilaceae), 키리스텐세넬라세에 (Christensenellaceae), 타네렐라세에 (Tannerellaceae), 락토바실라세에 (Lactobacillaceae), 펩토스트렙토코카세에 (Peptostreptococcaceae), 비피도박테리아세에 (Bifidobacteriaceae), 데플루비탈레세에 (Defluviitaleaceae), 코리오박테리아세에 (Coriobacteriaceae), 마이크로코카세에 (Micrococcaceae), 루미노코카세에 (Ruminococcaceae), 오실로스피라세에 (Oscillospiraceae), 라크노스피라세에 (Lachnospiraceae), 수테렐라세에 (Sutterellaceae), 엑티노미세타세에 (Actinomycetaceae), 에게텔라세에 (Eggerthellaceae), 펩토스트렙토코칼레스-티시에렐랄레스 (Peptostreptococcales-Tissierellales), 아케르만시아세에 (Akkermansiaceae) 과(Family)에 속하는 속(Genus)에서 선택되는 1종 이상의 미생물의 함량을 포함하는 것인, 진단 장치.

청구항 10

제 9 항에 있어서,

상기 미생물 데이터는 상기 혼합물이 혐기 조건에서 18시간 내지 24시간 동안 배양된 배양물에 포함된 내독소(endotoxin), 황화수소(hydrogen sulfide), 단쇄지방산(Short-chain fatty acids, SCFAs) 및 장균총 유래 대사체 중 하나 이상의 함량, 농도, 종류, 장균총에 포함된 균의 종류, 농도, 함량 또는 다양성 변화 중 적어도 하나를 포함하는 것인, 진단 장치.

청구항 11

제 9 항에 있어서,

상기 변수 선택 알고리즘은 보루타(Boruta) 알고리즘, 재귀 변수 제거(RFE: Recursive Feature Elimination) 알고리즘 중 적어도 하나를 포함하는 것인, 진단 장치.

청구항 12

제 9 항에 있어서,

상기 머신러닝 모델은 로지스틱회귀(Logistic Regression) 모델, GLMNET(Generalized linear) 모델, 랜덤포레스트(Random Forest) 모델, 그래디언트 부스팅(Gradient Boosting) 모델, XGB(Extreme Gradient Boosting) 모델 중 적어도 하나를 포함하는 것인, 진단 장치.

청구항 13

제 9 항에 있어서,

상기 미생물 관련 변수는 클로스트리디엄 (Clostridium), 서브도리그라눌럼 (Subdoligranulum), 알리스티페스 (Alistipes), 플라보니프랙터 (Flavonifractor), 블라우티아 (Blautia), 우에이셀라 (Weissella), 수테렐라 (Sutterella), 파라박테로이데스 (Parabacteroides), 락토바실러스 (Lactobacillus), 롬보웃시아 (Romboutsia), 인테스티니박터 (Intestinibacter), 루미노코커스 (Ruminococcus), 비피도박테리엄 (Bifidobacterium), 안에어로스티페스 (Anaerostipes), 도레아 (Dorea), 데플루비이탈레세에 (Defluviitaleaceae_UCG-011), 엑티노마이세스 (Actinomyces), 유박테리엄 (Eubacterium), 에게텔라 (Eggerthella), 류코노스톡 (Leuconostoc), 라크노스피라세에 (Lachnospiraceae_FCS020_group), 라크노스피라 (Lachnospira), 안에어로코커스 (Anaerococcus), 펩토니필러스 (Peptoniphilus), 테리스포로박터 (Terrisporobacter), 셀리모나스 (Sellimonas), 슬라키아 (Slackia), UCG-003, 캔디다터스_솔레아페리아 (Candidatus_Soleaferrea), 안에어로스티그넘 (Anaerostignum), 박테로이데스 (Bacteroides), UCG-002, 엔테로코커스 (Enterococcus), 투리시박터 (Turicibacter), 푸시카테니박터 (Fusicatenibacter), 아케르만시아 (Akkermansia), 튜리시박터 (Turicibacter) 속(Genus)에 속하는 종(Species)에서 선택되는 1종 이상의 미생물의 함량을 포함하는 것인, 진단 장치.

청구항 14

제 9 항에 있어서,

상기 미생물 관련 변수는 페르프링겐스 (Perfringens), 스하히이 (Shahii), 디스타소니스 (Distasonis), 포르미시게네란스 (Formicigenerans), 테타이오타오미크론 (Thetaiotaomicron), 피네골디이 (Finegoldii), 메르데 (Merdae), 온데르동키이 (Onderdonkii), 부티리쿰 (Butyricum), 롱검 (Longum), 문드티이 (Mundtii), 파라푸트리피쿰 (Paraputrificum), 락티스 (Lactis), 애니말리스 (Animalis), 무시니필라 (Muciniphila) 종(Species)에서 선택되는 1종 이상의 미생물의 함량을 포함하는 것인, 진단 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 머신러닝 모델을 이용하여 복부팽만 유무를 판별하는 방법 및 진단 장치에 관한 것이다.

배 경 기 술

[0002] 복부팽만은 소화불량의 한 증상으로서 배가 팽창되는 증상을 말한다. 복부팽만이 있는 경우 배가 단단해지거나, 더부룩한 느낌을 받거나, 체중이 늘거나, 신경이 예민해지는 등의 증상이 주로 동반된다.

[0003] 복부팽만은 전체 인구의 약 10~30%가 경험할 정도로 흔한 증상에 해당한다. 최근에는 서구화된 식생활, 폭식, 과식, 활동량 감소, 비만 등을 주된 원인으로 하여 장 속에 골고루 분포되어 있는 가스, 위액, 장액, 분변의 양이 많아져 복부 둘레가 늘어나며 발생한다.

[0004] 한양대학교구리병원에 의하면, 복부팽만은 단순히 복부가 불편한 증상이 아니라, 복통과 변비가 함께 동반되므로 삶의 질이 떨어질 수 있는 증상에 해당한다. 더 나아가 복부팽만은 기능성 위장장애 환자의 50%, 과민성대장증후군 환자의 23~96%, 만성변비환자의 56%가 겪는 증상으로, 복부팽만이 있다면 다른 질환의 존재도 의심할 수 있다.

[0005] 따라서 복부팽만 유무의 판단은 삶의 질 개선에도 도움이 될뿐더러 잠재적인 질병의 유무도 파악에 큰 도움이 된다.

[0006] 한편, 게놈(genome)은 염색체에 담긴 유전자를 말하고, 장균총(microbiota)은 미생물균총으로 환경 내 미생물 군집을 말하며, 마이크로바이옴(microbiome)은 환경 내 총 미생물 군집의 유전체를 말한다. 여기서, 마이크로바이옴 (microbiome)은 게놈(genome)과 장균총 (microbiota)이 합쳐진 것을 의미할 수 있다.

[0007] 최근, 이러한 장균총의 메타게놈 분석을 통해 복부팽만의 원인인자로 작용할 수 있는 미생물을 동정하여 복부팽만 유무를 진단하고자 하는 시도가 있다.

[0008] 이와 관련하여, 선행기술인 등록특허공보 제10-2057047호는 질병 예측 장치 및 이를 이용한 질병 예측 방법에 관한 것으로서, 특정인의 바이오 시그널에서 추출된 특정인 벡터를 학습 벡터와 비교하여 특정인의 질병을 예측하는 질병 예측 방법을 개시하고 있다.

[0009] 그러나, 선행기술에서는, 샘플을 배양 등 특별한 과정을 거치지 않고 세균 메타게놈 분석을 수행하는바, 각 피검체의 샘플들 간의 편차(bias)가 커 정확한 복부팽만의 원인인자를 도출하기 어렵다.

[0010] 또한, 처리되지 않은 각 피검체의 샘플들을 학습 데이터로서 머신러닝 모델을 학습시킬 경우, 학습 데이터에 노이즈가 많아 머신러닝 모델의 성능이 현저히 낮아지는 문제점이 있었다.

발명의 내용

해결하려는 과제

[0011] 본 발명은 상술한 문제점을 해결하기 위한 것으로서, 시료를 장내 환경 유사 조성물과 혼합한 혼합물의 분석 결과에 기초하여 복수의 미생물 데이터를 대상으로 미생물 관련 변수를 선택함으로써 복부팽만 유무를 진단하는 머신러닝 모델의 성능을 향상시키고자 한다.

[0012] 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제들로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0013] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 일 실시예는 머신러닝 모델을 이용하여 복부팽만 유무를 판별하는 방법은 개체로부터 채취한 장내 유래 물질을 장내 환경 유사 조성물과 혼합한 혼합물을 분석하는 단계, 상기 혼합물의 분석 결과에 기초하여 복수의 미생물 데이터를 추출하는 단계, 기설정된 변수 선택 알고리즘에 기초하여 상기 복수의 미생물 데이터 중 머신러닝 모델에 사용될 미생물 관련 변수를 선택하는 단계, 상기 미생물 관련 변수를 이용하여 상기 머신러닝 모델을 학습시키는 단계 및 검사 대상 객체로부터 채취한 미생물 데이터를 상기 학습된 머신러닝 모델에 입력하여 복부팽만 유무를 판별하는 단계를 포함할 수 있다. 상기 미생물 관련 변수는 클로스트리디아세에 (Clostridiaceae), 리케넬라세에 (Rikenellaceae), 덴테로코카세에 (Enterococcaceae), 박테로이다세에 (Bacteroidaceae), 데리시펠로트리카세에 (Erysipelotrichaceae), 류코노스토카세에 (Leuconostocaceae), 마리니필라세에 (Marinifilaceae), 키리스텐세넬라세에 (Christensenellaceae), 타네렐라세에 (Tannerellaceae), 락토바실라세에 (Lactobacillaceae), 펩토스트렙토코카세에 (Peptostreptococcaceae), 비피도박테리아세에 (Bifidobacteriaceae), 데플루비탈레세에 (Defluviitaleaceae), 코리오박테리아세에 (Coriobacteriaceae), 마이크로코카세에 (Micrococcaceae), 루미노코카세에 (Ruminococcaceae), 오실로스피라세에 (Oscillospiraceae), 라크노스피라세에 (Lachnospiraceae), 수테렐라세에 (Sutterellaceae), 엑티노미세타세에 (Actinomycetaceae), 에게텔라세에 (Eggerthellaceae), 펩토스트렙토코칼레스-티시에렐랄레스 (Peptostreptococcales-Tissierellales), 아케르만시아세에 (Akkermansiaceae) 과(Family)에 속하는 속(Genus)에서 선택되는 1종 이상의 미생물의 함량을 포함할 수 있다.

[0014] 또한, 본 발명의 다른 실시예는 머신러닝 모델을 이용하여 복부팽만 유무를 진단하는 장치는 개체로부터 채취한 장내 유래 물질을 장내 환경 유사 조성물과 혼합한 혼합물의 분석 결과에 기초하여 복수의 미생물 데이터를 추출하는 미생물 데이터 추출부, 기설정된 변수 선택 알고리즘에 기초하여 상기 복수의 미생물 데이터 중 머신러닝 모델에 사용될 미생물 관련 변수를 선택하는 변수 선택부, 상기 미생물 관련 변수를 이용하여 상기 머신러닝 모델을 학습시키는 학습부 및 검사 대상 객체로부터 채취한 미생물 데이터를 상기 학습된 머신러닝 모델에 입력하여 복부팽만 유무를 진단하는 진단부를 포함할 수 있다. 상기 미생물 관련 변수는 클로스트리디아세에 (Clostridiaceae), 리케넬라세에 (Rikenellaceae), 덴테로코카세에 (Enterococcaceae), 박테로이다세에 (Bacteroidaceae), 데리시펠로트리카세에 (Erysipelotrichaceae), 류코노스토카세에 (Leuconostocaceae), 마리니필라세에 (Marinifilaceae), 키리스텐세넬라세에 (Christensenellaceae), 타네렐라세에 (Tannerellaceae), 락토바실라세에 (Lactobacillaceae), 펩토스트렙토코카세에 (Peptostreptococcaceae), 비피도박테리아세에 (Bifidobacteriaceae), 데플루비탈레세에 (Defluviitaleaceae), 코리오박테리아세에 (Coriobacteriaceae), 마이크로코카세에 (Micrococcaceae), 루미노코카세에 (Ruminococcaceae), 오실로스피라세에 (Oscillospiraceae),

라크노스피라세에 (Lachnospiraceae), 수테렐라세에 (Sutterellaceae), 엑티노미세타세에 (Actinomycetaceae), 에게텔라세에 (Eggerthellaceae), 펩토스트렙토코칼레스-티시에렐랄레스 (Peptostreptococcales-Tissierellales), 아케르만시아세에 (Akkermansiaceae) 과(Family)에 속하는 속(Genus)에서 선택되는 1종 이상의 미생물의 함량을 포함할 수 있다.

[0015] 상술한 과제 해결 수단은 단지 예시적인 것으로서, 본 발명을 제한하려는 의도로 해석되지 않아야 한다. 상술한 예시적인 실시예 외에도, 도면 및 발명의 상세한 설명에 기재된 추가적인 실시예가 존재할 수 있다.

발명의 효과

[0016] 전술한 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 시료를 장내 환경 유사 조성물과 혼합한 혼합물의 분석 결과에 기초하여 복수의 미생물 데이터를 대상으로 미생물 관련 변수를 선택함으로써 복부팽만 유무를 진단하는 머신러닝 모델의 성능을 향상시킬 수 있다.

도면의 간단한 설명

[0017] 도 1은 본 발명의 일 실시예에 따른 진단 장치의 블록도를 도시한 도면이다.

도 2는 본 발명의 일 실시예에 따른 MCMOD기법을 나타낸 도면이다.

도 3은 본 발명의 일 실시예에 따른 MCMOD 기법을 통한 샘플 분석을 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 MCMOD 기법을 통한 샘플 분석 결과를 해석하는 것을 설명하기 위한 도면이다.

도 5는 본 발명의 일 실시예에 따라 선택된 미생물 관련 변수의 중요도를 나타낸 도면이다.

도 6은 본 발병의 일 실시예에 따라 선택된 미생물 관련 변수의 분류학 정보를 나타낸 도면이다.

도 7은 본 발명의 일 실시예에 따른 복부팽만 유무 판별 방법과 비교예의 방법에 따른 각 샘플의 분석 결과를 비교한 도면이다.

도 8은 본 발명의 일 실시예에 따른 복부팽만 유무 판별 방법과 비교예의 방법에 따른 각 샘플의 분석 결과를 비교한 도면이다.

도 9 내지 11은 본 발명의 실시예에 따라 선택된 미생물 관련 변수 별로 정상군과 질병군 간의 미생물 검출량 차이를 나타낸 도면이다.

도 12는 본 발명의 실시예에 따른 복부팽만 유무 판별 방법과 비교예의 방법에 따른 머신러닝 모델의 성능을 비교한 도면이다.

도 13은 본 발명의 일 실시예에 따른 복부팽만 유무 판별 방법과 비교예의 방법의 변수의 수에 따른 머신러닝 모델의 성능 변화를 도시한 도면이다.

도 14는 본 발명의 일 실시예에 따른 복부팽만 유무 판별 방법과 비교예의 방법에 따른 랜덤 포레스트 모델의 성능을 비교한 도면이다.

도 15는 본 발명의 일 실시예에 따른 복부팽만 유무 판별 방법과 비교예의 방법에 따른 XGB 모델의 성능을 비교한 도면이다.

도 16은 본 발명의 일 실시예에 따른 복부팽만 유무 판별 방법을 도시한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0018] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0019] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아

니라 다른 구성요소를 더 포함할 수 있는 것을 의미하며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0020] 본 명세서에 있어서 '부(部)'란, 하드웨어에 의해 실현되는 유닛(unit), 소프트웨어에 의해 실현되는 유닛, 양방을 이용하여 실현되는 유닛을 포함한다. 또한, 1 개의 유닛이 2 개 이상의 하드웨어를 이용하여 실현되어도 되고, 2 개 이상의 유닛이 1 개의 하드웨어에 의해 실현되어도 된다.

[0021] 본 명세서에 있어서 단말 또는 디바이스가 수행하는 것으로 기술된 동작이나 기능 중 일부는 해당 단말 또는 디바이스와 연결된 서버에서 대신 수행될 수도 있다. 이와 마찬가지로, 서버가 수행하는 것으로 기술된 동작이나 기능 중 일부도 해당 서버와 연결된 단말 또는 디바이스에서 수행될 수도 있다.

[0022] 이하 첨부된 도면을 참고하여 본 발명의 일 실시예를 상세히 설명하기로 한다.

[0023] 도 1은 본 발명의 일 실시예에 따른 진단 장치의 블록도를 도시한 도면이다. 도 1을 참조하면, 진단 장치(1)는 미생물 데이터 추출부(100), 변수 선택부(110), 학습부(120) 및 진단부(130)를 포함할 수 있다. 본원에 있어서 진단 장치(1)는 복부팽만 유무를 판별하는 판별 장치일 수 있다.

[0024] 진단 장치(1)의 일례는 데스크탑, 노트북 등과 같은 퍼스널 컴퓨터(personal computer)뿐만 아니라 유무선 통신이 가능한 모바일 단말을 포함할 수 있다. 모바일 단말은 휴대성과 이동성이 보장되는 무선 통신 장치로서, 스마트폰(smartphone), 태블릿 PC, 웨어러블 디바이스뿐만 아니라, 블루투스(BLE, Bluetooth Low Energy), NFC, RFID, 초음파(Ultrasonic), 적외선, 와이파이(WiFi), 라이파이(LiFi) 등의 통신 모듈을 탑재한 각종 디바이스를 포함할 수 있다. 다만, 진단 장치(1)는 도 1에 도시된 형태 또는 앞서 예시된 것들로 한정 해석되는 것은 아니다.

[0025] 진단 장치(1)는 개체로부터 채취한 시료에서 장내 환경 이상에 의한 복부팽만 유무를 진단하기 위한 바이오마커를 검출할 수 있다.

[0026] 예를 들어, 진단 장치(1)는 샘플 준비 과정, 샘플 전처리 과정, 샘플 분석 과정 및 데이터 분석 과정, 도출된 데이터를 토대로 복부팽만 유무를 진단할 수 있다. 본원에 있어서, "진단"이란 머신러닝 모델의 출력값을 통해 복부팽만 유무를 판별 또는 예측하는 것은 의미할 수 있다.

[0027] 일례에 있어서, 바이오마커는 장내에서 검출되는 물질일 수 있으며, 구체적으로, 장균총, 내독소, 황화수소, 장내 미생물 대사체, 단쇄지방산 등을 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다.

[0028] 미생물 데이터 추출부(100)는 개체로부터 채취한 시료를 장내 환경 유사 조성물과 혼합한 혼합물의 분석 결과에 기초하여 복수의 미생물 데이터를 추출할 수 있다. 여기서, 복수의 미생물 데이터는 학습에 사용될 훈련 데이터(Training set) 및 테스트 데이터(Test set)로 분류될 수 있고, 분류의 비율은 9:1, 7:3, 5:5 등으로 다양할 수 있고, 바람직하게는 7:3 비율로 분류될 수 있다.

[0029] 본 발명에 따르면, 시료를 장내 환경 유사 조성물과 혼합한 혼합물을 분석하는 전처리가 수행된다. 본원에 있어서, 전처리는 MCMOD(Meta-culture Multi-Omics Diagnose)라고 명명될 수 있다.

[0030] 예를 들어, 체내 장내 미생물 환경을 가장 용이하게 대표할 수 있는 사람 및 다양한 동물의 분변 샘플을 대상으로 체외(in-vitro)에서 분변 유래 마이크로바이옴(microbiome)과 대사물질(metabolites) 분석이 수행된다.

[0031] 여기서, "개체"는 장내 환경에 이상이 있거나, 장내 환경 이상에 의한 질병이 발병 또는 발병할 가능성이 있거나, 또는 장내 환경이 개선되어야 할 필요성이 있는 모든 생물체를 의미하며, 구체적인 예로, 마우스, 원숭이, 소, 돼지, 미니돼지, 가축, 인간 등을 포함하는 포유동물, 조류, 양식어류 등을 제한 없이 포함할 수 있다.

[0032] "시료"는 상기 개체로부터 유래한 물질을 의미하며, 예를 들어 장내 유래 물질일 수 있다.

[0033] “시료”는 구체적으로 세포, 소변, 분변 등일 수 있으나, 장균총, 장내 미생물 대사체, 내독소, 단쇄지방산 등 장내에 존재하는 물질을 검출할 수 있는 한, 그 종류가 이에 제한되는 것은 아니다.

[0034] “장내 환경 유사 조성물”은 상기 개체의 장내 환경을 체외(in vitro) 에서 동일 또는 유사하게 형성(mimicking)하기 위한 조성물일 수 있다. 예를 들어, 장내 환경 유사 조성물은 배양 배지 조성물일 수 있으나, 이에 제한되는 것은 아니다.

[0035] 장내 환경 유사 조성물은 L-시스테인 염산염(L-cystein Hydrochloride) 및 뮤신(Mucin)을 포함할 수 있다.

[0036] 여기서, "L-시스테인 염산염(L-cystein Hydrochloride)"은 아미노산류 강화제 중 하나로서, 생체 내에서 글루타치온의 구성성분으로 대사에 중요한 역할을 하며, 과일주수 등의 갈변 방지 및 비타민 C의 산화 방지 등에도 이용된다.

[0037] L-시스테인 염산염은 예를 들어, 0.001%(w/v) 내지 5%(w/v)의 농도로 포함되는 것일 수 있으며, 구체적으로 0.01%(w/v) 내지 0.1%(w/v)의 농도로 포함될 수 있다.

[0038] L-시스테인 염산염은 다양한 L-시스테인의 제형 또는 형태 중 하나로서, 상기 조성물은 L-시스테인 뿐만 아니라, 다른 형태의 염이 포함된 L-시스테인을 포함할 수 있다.

[0039] 뮤신(Mucin)"은 점막에서 분비되는 점액물질로 점액소 또는 점소라고도 불리우며, 턱밑샘 뮤신이 있으며 그 외에 위점막뮤신, 소장뮤신 등이 있다. 뮤신은 당단백질의 일종으로서, 실제 장 내 미생물들이 활용할 수 있는 탄소원 및 질소원이 되는 에너지원 중 하나라고 알려져 있다.

[0040] 뮤신은 예를 들어, 0.01%(w/v) 내지 5%(w/v)의 농도로 포함되는 것일 수 있으며, 구체적으로 0.1%(w/v) 내지 1%(w/v)의 농도로 포함되는 것일 수 있으나, 이에 제한되는 것은 아니다.

[0041] 일례에 있어서, 장내 환경 유사 조성물은 뮤신을 제외한 영양물질을 포함하지 않을 수 있으며, 구체적으로 단백질 및 탄수화물과 같은 질소원 및/또는 탄소원을 포함하지 않는 것을 특징으로 하는 것일 수 있다.

[0042] 탄소원 및 질소원이 되는 단백질은 트립톤, 펩톤 및 효모 추출물 중 하나 이상일 수 있으나, 이에 제한되는 것은 아니며, 구체적으로 트립톤일 수 있다.

[0043] 탄소원이 되는 탄수화물은 글루코스, 프럭토스, 갈락토스와 같은 단당류와 말토오스, 락토오스와 같은 이당류 중 하나 이상일 수 있으나, 이에 제한되는 것은 아니며, 구체적으로 글루코스일 수 있다.

[0044] 일례에 있어서, 장내 환경 유사 조성물은 글루코스(Glucose) 및 트립톤(Tryptone)을 포함하는 않는 것일 수 있으나, 이에 제한되는 것은 아니다.

[0045] 장내 환경 유사 조성물은 염화나트륨(NaCl), 탄산나트륨(NaHCO3), KCl(염화칼륨) 및 헤민(Hemin)으로 이루어진 군에서 선택된 하나 이상을 추가로 포함할 수 있으며, 염화나트륨은 예를 들어, 10 내지 100mM의 농도로 포함되는 것일 수 있고, 탄산나트륨은 예를 들어, 10 내지 100mM의 농도로 포함되는 것일 수 있고, 염화칼륨은 예를 들어, 1 내지 30mM의 농도로 포함되는 것일 수 있으며, 헤민은 예를 들어, 1x10-6 g/L 내지1x10-4 g/L 농도로 포함되는 것일 수 있으나, 이에 제한되는 것은 아니다.

[0046] 전처리에서, 혼합물을 혐기 조건에서 18시간 내지 24시간 동안 배양할 수 있다.

[0047] 예를 들어, 혐기 챔버 내에서 분변과 배지의 균질화된 혼합물을 96-웰 플레이트 등의 배양 플레이트에 각각 동일 양씩 분주한다. 여기서, 배양은 12시간 내지 48시간동안 수행하는 것일 수 있으며, 구체적으로 18시간 내지 24시간동안 수행하는 것일 수 있으나, 이에 제한되는 것은 아니다.

[0048] 이어서, 온도, 습도 및 모션을 장내 환경과 유사하게 형성한 채로 혐기 조건에서 플레이트를 배양하여 각 실험군을 발효 배양시킨다.

[0049] 혼합물의 배양 후, 혼합물이 배양된 배양물을 분석한다. 배양물의 분석은 예를 들어, 배양물에 포함된 내독소(endotoxin), 황화수소(hydrogen sulfide), 단쇄지방산(Short-chain fatty acids, SCFAs) 및 장균총 유래 대사체 중 하나 이상의 함량, 농도, 종류, 장균총에 포함된 균의 종류, 농도, 함량 또는 다양성 변화 중 적어도 하나를 포함하는 미생물 데이터를 추출하는 것일 수 있으나, 이에 제한되는 것은 아니다.

[0050] 여기서, "내독소(endotoxin)"는 세균의 세포 내부에서 발견되는 독성 물질로 단백질·다당류·지질의 복합체로 이루어진 항원 등이다. 일례에 있어서, 내독소는 LPS(Lipopolysaccharide)를 포함하는 것일 수 있으나, 이에 제한되는 것은 아니며, 상기 LPS는 구체적으로 그람 음성(Gram negative), 프로 염증성(Pro-inflammatory)일 수 있다.

[0051] "단쇄지방산 (short-chain fatty acid : SCFA)"은 단쇄지방산은 탄소수가 6개 이하인 짧은 길이의 지방산을 의미하는 것으로서, 장내 미생물로부터 생성되는 대표적인 대사산물이다. 단쇄지방산은 면역력 증가, 장내 림프구 안정, 인슐린 신호 저하, 교감 신경 자극 등 체내에 유용한 기능을 가지고 있다.

[0052] 일례에 있어서, 단쇄지방산은 포름산(Formate), 아세트산(Acetate), 프로피온산(Propionate), 뷰티르산(Butyrate), 아이소뷰티르산 (Isobutyrate), 발레르산(Valerate) 및 아이소발레르산(Iso-valerate)으로 이루어

진 군에서 선택되는 하나 이상을 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다.

[0053] 배양물의 분석 방법은 흡광도 분석법, 크로마토그래피 분석법, 차세대시퀀싱방법(Next Generation Sequencing) 등의 유전자 분석법, 메타지놈 분석법 등의 통상의 기술자가 상기 분석을 위해 이용할 수 있는 다양한 분석법을 이용할 수 있다.

[0054] 배양물의 분석에 있어서, 배양물을 원심 분리하여 상등액과 침전물을 분리한 후, 상기 상등액 및 상기 침전물(pallet)을 분석할 수 있다. 예를 들어, 상등액으로부터 대사체, 단쇄지방산, 독성 물질 등을 분석하고, 침전물로부터 장균총 분석을 수행할 수 있다.

[0055] 예를 들어, 배양이 종료된 후, 배양된 각각의 실험군을 원심분리하여 얻어진 상등액으로부터 흡광도 측정분석법과 크로마토그래피 분석법을 통해 황화수소 및 박테리아 LPS(내독소) 등의 독성 물질 분석 및 단쇄지방산 등의 미생물 대사체 분석을 수행하며, 원심분리하여 얻어진 침전물(pallet)으로부터 배양-비의존적장균총 분석(Culture-independent analysis method)을 수행한다. 예를 들어, N,N-디메틸-p-페닐렌디아민 (N,N-dimethyl-p-phenylene-diamine)과 염화철 (FeCl3)로 반응시키는 메틸렌블루법(methylene blue method)을 통해서 배양을 통해 생성된 황화수소의 변화량을 측정하고, 내독소 어세이 키트 (Endotoxin assay kit) 분석을 통해 염증반응 증진요인 중 하나인 내독소(Endotoxin)의 레벨을 측정할 수 있다. 또한 가스 크로마토그래피 분석법을 활용하여 미생물 대사체인 아세테이트, 프로피오네이트, 부티레이트 등의 단쇄지방산을 분석할 수 있다.

[0056] 장균총은 시료 내 유전체를 전부 추출한 후 GULDA방법에서 제시된 박테리아 특이적인 프라이머를 사용한 실시간 PCR 분석법(real-time PCR)이나 차세대시퀀싱(Next Generation Sequencing)과 같은 메타지놈(metagenome)분석을 통하여 유전체 기반의 분석법으로 분석할 수 있다.

[0057] 본 발명에 따르면, 장내 환경 유사 조성물을 통해 체외에서 장내 환경을 구현한 상태에서 배양물을 분석함으로써 머신러닝 전에 학습 데이터를 최적화하여 학습 데이터 간의 편차를 줄일 수 있다.

[0058] 이에 따라, 후술하는 미생물 관련 변수의 선택을 용이하게 하고, 또한 이러한 미생물 관련 변수를 통해 머신러닝 모델을 학습함으로써 머신러닝 모델의 성능을 향상시킬 수 있다. 따라서, 학습된 머신러닝 모델을 통해 복부팽만 유무 진단의 정확도를 높일 수 있다.

[0059] 변수 선택부(110)는 기설정된 변수 선택 알고리즘에 기초하여 복수의 미생물 데이터 중 미생물 관련 변수를 머신러닝 모델에 사용될 변수로서 선택(즉, Feature Selection)할 수 있다. 미생물 관련 변수의 수는 108개 이상일 수 있다.

[0060] 머신러닝 모델을 생성함에 있어서 변수(features, 또는 variables, attributes)가 사용되는데, 많은 수의 변수 또는 부적절한 변수들이 사용되면 머신러닝 모델이 과적합(Overfitting)되거나 예측 정확도가 감소하는 문제가 발생한다.

[0061] 이에, 머신러닝 모델이 높은 예측 정확도를 갖기 위해서는 적절한 변수들의 조합을 사용할 필요가 있다. 즉, 예측하고자 하는 반응변수와 가장 연관성이 높은 변수들을 선택하여 가능한 한 적은 수의 변수를 사용하면서 머신러닝 모델의 복잡도(complexity)를 낮출 수 있다.

[0062] 변수 선택 알고리즘은 예를 들어, 보루타(Boruta) 알고리즘, 재귀 변수 제거(RFE: Recursive Feature Elimination) 알고리즘 중 적어도 하나를 포함할 일 수 있다.

[0063] 기설정된 변수 선택 알고리즘으로부터 선택된 미생물 관련 변수는 클로스트리디아세에 (Clostridiaceae), 리켄넬라세에 (Rikenellaceae), 덴테로코카세에 (Enterococcaceae), 박테로이다세에 (Bacteroidaceae), 데리시펠로트리카세에 (Erysipelotrichaceae), 류코노스토카세에 (Leuconostocaceae), 마리니필라세에 (Marinifilaceae), 키리스텐세넬라세에 (Christensenellaceae), 타네렐라세에 (Tannerellaceae), 락토바실라세에 (Lactobacillaceae), 펩토스트렙토코카세에 (Peptostreptococcaceae), 비피도박테리아세에 (Bifidobacteriaceae), 데플루비탈레세에 (Defluviitaleaceae), 코리오박테리아세에 (Coriobacteriaceae), 마이크로코카세에 (Micrococcaceae), 루미노코카세에 (Ruminococcaceae), 오실로스피라세에 (Oscillospiraceae), 라크노스피라세에 (Lachnospiraceae), 수테렐라세에 (Sutterellaceae), 엑티노미세타세에 (Actinomycetaceae), 에게텔라세에 (Eggerthellaceae), 펩토스트렙토코칼레스-티시에렐랄레스 (Peptostreptococcales-Tissierellales), 아케르만시아세에 (Akkermansiaceae) 과(Family)에 속하는 속(Genus)에서 선택되는 1종 이상의 미생물의 함량을 포함할 수 있다.

[0064] 일례에 있어서, 기설정된 변수 선택 알고리즘으로부터 선택된 미생물 관련 변수는 예를 들어, 클로스트리디엄

(Clostridium), 서브도리그라눌럼 (Subdoligranulum), 알리스티페스 (Alistipes), 플라보니프랙터 (Flavonifractor), 블라우티아 (Blautia), 우에이셀라 (Weissella), 수테렐라 (Sutterella), 파라박테로이데스 (Parabacteroides), 락토바실러스 (Lactobacillus), 롬보웃시아 (Romboutsia), 인테스티니박터 (Intestinibacter), 루미노코커스 (Ruminococcus), 비피도박테리엄 (Bifidobacterium), 안에어로스티페스 (Anaerostipes), 도레아 (Dorea), 데플루비이탈레세에 (Defluviitaleaceae_UCG-011), 엑티노마이세스 (Actinomyces), 유박테리엄 (Eubacterium), 에게텔라 (Eggerthella), 류코노스톡 (Leuconostoc), 라크노스피라세에 (Lachnospiraceae_FCS020_group), 라크노스피라 (Lachnospira), 안에어로코커스 (Anaerococcus), 펩토니필러스 (Peptoniphilus), 테리스포로박터 (Terrisporobacter), 셀리모나스 (Sellimonas), 슬라키아 (Slackia), UCG-003, 캔디다터스_솔레아페리아 (Candidatus_Soleaferrea), 안에어로스티그넘 (Anaerostignum), 박테로이데스 (Bacteroides), UCG-002, 엔테로코커스 (Enterococcus), 투리시박터 (Turicibacter), 푸시카테니박터 (Fusicatenibacter), 아케르만시아 (Akkermansia), 튜리시박터 (Turicibacter) 속(Genus)에 속하는 종(Species)에서 선택되는 1종 이상의 미생물의 함량을 더 포함할 수 있다.

[0065] 일례에 있어서, 기설정된 변수 선택 알고리즘으로부터 선택된 미생물 관련 변수는 예를 들어, 페르프링겐스 (Perfringens), 스하히이 (Shahii), 디스타소니스 (Distasonis), 포르미시게네란스 (Formicigenerans), 테타이오타오미크론 (Thetaiotaomicron), 피네골디이 (Finegoldii), 메르데 (Merdae), 온데르동키이 (Onderdonkii), 부티리쿰 (Butyricum), 롱검 (Longum), 문드티이 (Mundtii), 파라푸트리피쿰 (Paraputrificum), 락티스 (Lactis), 애니말리스 (Animalis), 무시니필라 (Muciniphila) 종(Species)에서 선택되는 1종 이상의 미생물의 함량을 더 포함할 수 있다.

[0066] 학습부(120)는 미생물 관련 변수를 이용하여 머신러닝 모델을 학습시킬 수 있다.

[0067] 예를 들어, 학습부(120)는 미생물 데이터(학습 데이터)마다 복부팽만 유무에 관한 라벨링 및 선택된 변수에 관한 미생물의 함량에 기초하여 지도 학습을 수행하여 미생물 데이터마다 복부팽만 유무를 예측하도록 머신러닝 모델을 학습시킬 수 있다.

[0068] 머신러닝 모델은 예를 들어 로지스틱 회귀(Logistic Regression) 모델, GLMNET(Generalized linear) 모델, 랜덤포레스트(Random Forest) 모델, 그래디언트 부스팅(Gradient Boosting) 모델, XGB(Extreme Gradient Boosting) 모델 중 적어도 하나를 포함할 수 있다.

[0069] 진단부(130)는 검사 대상 객체로부터 채취한 미생물 데이터를 학습된 머신러닝 모델에 입력하여 복부팽만 유무를 진단할 수 있다.

[0070] 예를 들어, 진단부(130)는 머신러닝 모델의 출력값인 복부팽만 유무에 기초하여 복부팽만을 진단할 수 있다. 즉, 진단부(130)는 머신러닝 모델의 출력값에 기초하여 검사 대상 객체의 복부팽만 유무를 판별하거나 검사 대상 객체의 복부팽만 발생 확률을 예측할 수 있다.

[0071] 이하, 본원의 실시예를 상세히 설명한다. 그러나, 본원은 이에 제한되지 않는다.

[0073] **[실시예]**

[0074] **실시예 1. MCMOD 처리 또는 미처리 후 재귀 변수 제거 알고리즘에 기초하여 선택된 미생물 관련 변수**

[0075] 실시예 1의 MCMOD 처리 또는 미처리 후 재귀 변수 제거 알고리즘에 기초하여 선택된 미생물 관련 변수를 확인하기 위해, 하기와 같은 실험을 수행하였다.

[0076] 본 발명에 따르면, 시료를 장내 환경 유사 조성물과 혼합한 혼합물을 분석하는 전처리가 수행된다. 본원에 있어서, 상술한 전처리는 MCMOD라고 명명될 수 있다. 한편, 본원에 있어서, 비교예는 시료에 상술한 전처리를 수행하지 않고 통상의 전처리만을 수행하여 추출된 미생물 데이터를 통한 복부팽만 유무 판별 방법에 관한 것이다. 이와 관련하여, 비교예를 위한 통상의 전처리를 SMOD라고 명명한다.

[0077] 시료는, 하기 표 1과 같이 33명의 복부팽만(Gasbloating) 환자(질병군)와 105명의 정상인(정상군)으로부터 받은 문진 자가 응답 결과를 바탕으로 한 간이임상 데이터 세트(분변)의 MCMOD 및 SMOD의 미생물 데이터를 사용하였고, 특히 클래스 불균형 해소를 위해 데이터 세트에 대해 오버샘플링(oversampling) 및 언더샘플링(undersampling)을 수행하여 해당 데이터 세트를 78개의 정상 데이터 및 78개의 복부팽만 데이터를 포함하는 총 156개의 데이터 세트로 변환하였다.

표 1

질병 및 문진 항목	분류	데이터 출처 (수집경로)	질병기준	원본 데이터 샘플수			원본 데이터						Oversampling					
							Train Set			Test Set			Train Set			Test Set		
				질병군	정상군	합계	질병군	정상군	합계	질병군	정상군	합계	질병군	정상군	합계	질병군	정상군	합계
복부팽만 (Gas bloating)	문진표	기름범원	본인 응답 (문진당시 상태)	33	105	138	26	78	104	7	27	34	78	78	156	-		

[0078]

[0079] 미생물 데이터는 7:3의 비율로 학습에 사용될 훈련 데이터(Train set)와 테스트 데이터(Test set)로 분류하였다.

[0080] 이후, 훈련 데이터를 대상으로 보루타 알고리즘과 재귀 변수 제거 알고리즘을 통해 변수 선택을 수행하여 머신러닝 모델에 사용될 미생물 관련 변수를 선택하였다. 한편, 테스트 데이터는 후술하는 바와 같이 머신러닝 모델의 성능 평가에 사용되었다.

[0081] 도 5는 본 발명의 일 실시예에 따른 선택된 미생물 관련 변수를 설명하기 위한 도면이다. 재귀 변수 제거 알고리즘을 통해 가장 정확도가 높은 변수 집단으로서, 실시예의 경우 108개, 비교예의 경우 109개의 미생물 관련 변수가 선택되었다.

[0082] 도 6은 본 발명의 일 실시예에 있어서 선택된 미생물 관련 변수의 분류학적 정보를 도시한다. 도 6에서, 축약명 앞 알파벳은 분류학적 위치를 의미한다. 즉, 'p' 는 문(Phylum), 'c' 는 강(Class), 'o' 는 목(Order, 'f' 는 과(Family), 'g' 는 속(Genus) 및 's' 는 종(Species)을 의미한다.

[0083] 예를 들어, MCMOD에 있어서, 선택된 복수의 미생물 관련 변수 중 정확도가 높은 미생물 관련 변수는 액티노박테리오타문 (Actinobacteriota)의 미생물 일 수 있다.

[0085] **비교예 1. MCMOD 처리한 분변 샘플과 MCMOD 처리하지 않은 분변 샘플의 분석 결과**

[0086] 한 사람의 분변을 8일 동안 채취하여, 날짜별 8개의 분변 샘플(J01, J02, J03, J04, J06, J08, J09, J10)을 MCMOD 처리하였고, MCMOD 처리한 분변 샘플을 차세대시퀀싱으로 미생물의 유전자를 분석했다(실시예). 마찬가지로, MCMOD 처리하지 않은 분변 샘플을 차세대시퀀싱으로 미생물의 유전자를 분석했다(비교예).

[0087] 도 7은 본 발명의 일 실시예에 따른 복부팽만 유무 판별 방법과 비교예의 방법에 따른 각 샘플의 분석 결과를 비교한 도면이고, 도 8은 본 발명의 일 실시예에 따른 복부팽만 유무 판별 방법과 비교예의 방법에 따른 각 샘플의 분석 결과를 비교한 도면이다.

[0088] 도 7의 (a)은 각 분변 샘플들의 베타 다양성을 Unweighted Unifrac Distance 를 사용하여 PCoA plot으로 표현한다. 도 7의 (a)의 PCoA plot에 도시된 바와 같이, MCMOD 처리한 분변 샘플들은 상대적으로 모여있는 형태를 띄는데 반해, MCMOD 처리하지 않은 분변 샘플들은 상대적으로 흩어져 있는 형태를 띄는 것을 확인할 수 있다.

[0089] 도 7의 (b)는 PCoA plot 상의 각 그룹(실시예 및 비교예) 내의 8개의 점 사이의 거리를 Box plot으로 표현한다.

[0090] Box plot에서 확인할 수 있듯이, 실시예의 경우, 분변 샘플 간의 차이가 비교예에 비해 통계적으로 유의미하게 적음을 확인할 수 있다.

[0091] 도 7의 (c)는 PCoA plot 상의 각 그룹(실시예 및 비교예) 내의 8개의 점 사이의 거리를 나타낸다.

[0092] 각 그룹에 8개의 샘플이 존재하므로, 각 그룹 내의 두 샘플 간 거리는 총 28가지의 종류를 갖는다. 이 28가지의 종류를 시간 순으로 샘플을 ${}_2C_2$ 부터 ${}_8C_2$ 까지 그룹화하였다.

[0093] J01 분변샘플이 가장 먼저 수집되었고 J10 분변샘플이 가장 나중에 수집되었으므로 ${}_2C_2$ (N=1) 그룹에서는 가장 먼저 수집된 분변 두 샘플간의 거리 (J01과 J02 샘플 간의 거리)를 구했다.

[0094] ${}_3C_2$ (N=3) 그룹에서는 그 다음으로 수집된 분변 샘플(J03)을 포함하여 3개 샘플들에서 각 샘플들 사이의 거리 (J01과J02, J01과J03, J02와J03)를 구하고 이 거리들의 평균과 표준오차를 표현했다.

[0095] ${}_4C_2$ (N=6) 그룹에서는 그 다음으로 수집된 분변 샘플(J04)을 포함하여 4개 샘플들에서 각 샘플들 사이의 거리 (J01과J02, J01과J03, J01과J04, J02와J03, J02와J04, J03과J4)를 구하고 이 거리들의 평균과 표준오차를 표현했다.

[0096] 마찬가지로, ${}_8C_2$ (N=28) 그룹에서는 마지막으로 수집된 분변 샘플(J10)을 포함하여 8개 샘플들에서 각 샘플들 사이의 거리(총 28가지)를 구하고 이 거리들의 평균과 표준오차를 표현했다.

[0097] PCoA plot에서의 거리 수치로 확인할 수 있듯이, 실시예에 따른 분변 샘플 그룹(${}_2C_2$ ~ ${}_8C_2$)의 샘플 간의 차이가 비교예에 비해 통계적으로 유의미하게 적음을 확인 할 수 있다.

[0098] 도 8은 두 그룹(실시예 및 비교예)에 대하여 PERMANOVA 분산을 분석한 결과를 나타낸다.

[0099] 도 8은 (b)와 같이 PERMANOVA 분산 분석 결과, Pr (> F) 값이 0.001 로 매우 작아 두 그룹(실시예 및 비교예)의 모평균이 동일하지 않음을 알 수 있다. 이는, 두 그룹이 통계적으로 유의하게 차이가 있다는 것을 나타낸다.

[0100] 또한, 각 그룹의 중심으로부터 각 분변 샘플의 평균거리(Average distance to median)가 비교예(0.2340)보다 실시예(0.1792)가 더 가까움을 확인할 수 있고, 이는 실시예가 비교예에 비해 노이즈가 적다는 것을 의미한다.

[0101] 살펴본 바와 같이, MCMOD 처리한 분변 샘플들의 경우, 분변 샘플들간의 편차(bias)가 적어 분변 샘플이 비교적 적은 노이즈(Noise)를 갖고 있고, 이에 따라 적은 변동성(Fluctuation)을 갖는다.

[0102] 즉, 본 발명에 따르면, 변수 선택 및 머신러닝 학습 전에 분변 샘플을 MCMOD 처리함으로써 변수 선택을 용이하게 하고, 또한 후술하는 바와 같이, 머신러닝 모델을 학습함으로써 머신러닝 모델의 성능을 향상시킬 수 있다.

[0104] **비교예 2. MCMOD 처리한 분변 샘플과 MCMOD 처리하지 않은 분변 샘플 각각으로부터 획득한 학습 데이터를 이용하여 학습한 머신러닝 모델의 성능 비교**

[0105] 실시예 1을 통해 채취한 분변 샘플을 MCMOD 처리하여 미생물 데이터를 추출하였고(실시예), MCMOD 처리하지 않고 미생물 데이터를 추출하였다(비교예).

[0106] 재귀 변수 제거 알고리즘을 통해 실시예의 경우, 미생물 데이터로부터 108개의 미생물 관련 변수를 선택하였고, 비교예의 경우, 미생물 데이터로부터 109개의 미생물 관련 변수를 선택하였다.

[0107] 실시예 및 비교예의 미생물 데이터 및 미생물 관련 변수를 이용하여 로지스틱 회귀 분석(LRA: Logistic Regression Analysis) 모델, 랜덤포레스트 (RF, Random Forest) 모델, Glmnet 모델, 그래디언트 부스팅(Gradient Boosting) 모델 및 XGB(Extreme Gradient Boost) 모델 각각을 학습시킨 후, 각 머신러닝 모델의 성능을 평가하였다.

[0108] 도 9 내지 11은 본 발명의 실시예에 따라 선택된 미생물 관련 변수 별, 정상군과 질병군 간의 미생물 검출량 차이를 나타낸 도면이다. 중요도 순위가 높은 미생물 변수의 경우, 정상군과 비교했을 때 질병군에서 상대적으로 검출량이 많은 것을 알 수 있다.

[0109] 도 12는 각 머신러닝 모델의 Roc curve와 AUC score를 나타낸다. 도 12에 나타난 바와 같이, 실시예의 미생물 데이터를 이용하여 머신러닝 모델을 학습한 경우, 비교예에 비해 모든 머신러닝 모델의 성능이 높은 것을 확인할 수 있다. 이때, 도 13에 도시된 바와 같이, 실시예의 경우 전체적으로 비교예에 비해 머신러닝 모델의 성능이 높게 유지되나, 그 중에서 108개 이상의 변수를 선택했을 때 성능 차이가 확연하게 드러남을 알 수 있다.

[0110] 도 14는 실시예의 미생물 데이터를 이용하여 학습된 랜덤 포레스트 모델과 비교예의 미생물 데이터를 이용하여 학습된 랜덤 포레스트 모델의 정확도, 민감도 및 특이도를 나타내며, 도 15는 실시예의 미생물 데이터를 이용하여 학습된 XGB 모델과 비교예의 미생물 데이터를 이용하여 학습된 XGB 모델의 정확도, 민감도 및 특이도를 나타낸다.

[0111] 여기서, No information rate는 테스트 세트에서 하나의 군으로(질병 또는 정상) 일괄적으로 예측했을 때의 정확도를 나타낸다. 예를 들어, 테스트 세트에서 질병군이 6명, 실험군이 4명이 존재할 때, 모든 테스트 세트를 질병군으로만 예측할 때의 No information rate 는 0.6 이다.

[0112] 도 14 및 15에 도시된 바와 같이, 실시예의 미생물 데이터를 이용하여 학습된 머신러닝 모델이 비교예의 미생물 데이터를 이용하여 학습된 머신러닝 모델보다 높은 정확도, 민감도 및 특이도를 갖는다는 것을 확인할 수 있다.

[0113] 도 16은 본 발명의 일 실시예에 따른 복부팽만 유무 판별 방법을 나타낸 흐름도이다. 도 16에 도시된 일 실시예에 따른 복부팽만 유무 판별 방법은 도 1에 도시된 진단 장치에서 시계열적으로 처리되는 단계들을 포함한다. 따라서, 이하 생략된 내용이라고 하더라도 도 16에 도시된 일 실시예에 따라 수행되는 복부팽만 유무 판별 방법에도 적용된다.

[0114] 도 16를 참조하면, 단계 S1600에서 개체로부터 채취한 장내 유래 물질을 장내 환경 유사 조성물과 혼합한 혼합물을 분석할 수 있다.

[0115] 단계 S1610에서 혼합물의 분석 결과에 기초하여 복수의 미생물 데이터를 추출할 수 있다.

[0116] 단계 S1620에서 기설정된 변수 선택 알고리즘에 기초하여 복수의 미생물 데이터 중 머신러닝 모델에 사용될 미생물 관련 변수를 선택할 수 있다.

[0117] 단계 S1630에서 미생물 관련 변수를 이용하여 머신러닝 모델을 학습시킬 수 있다.

[0118] 단계 S1640에서 미생물 관련 변수를 이용하여 머신러닝 모델을 학습시킬 수 있다.

[0119] 검사 대상 객체로부터 채취한 미생물 데이터를 학습된 머신러닝 모델에 입력하여 복부팽만 유무를 판별할 수 있다.

[0120] 도 16를 통해 설명된 복부팽만 유무 판별 방법은 매체에 저장된 컴퓨터 프로그램의 형태로 구현되거나, 컴퓨터

에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행 가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체를 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다.

[0121] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

[0122] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

[0123] 1: 진단 장치

100: 미생물 데이터 추출부

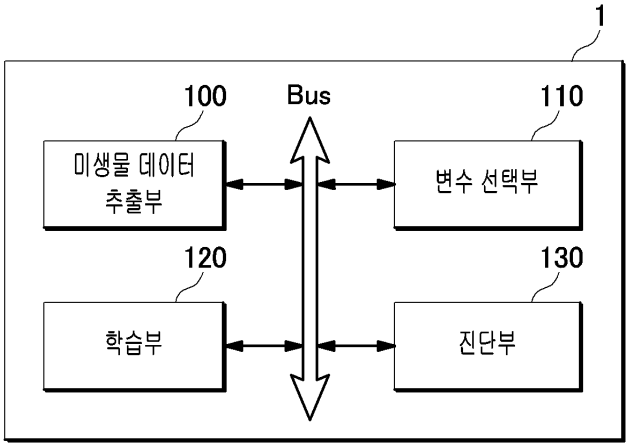
110: 변수 선택부

120: 학습부

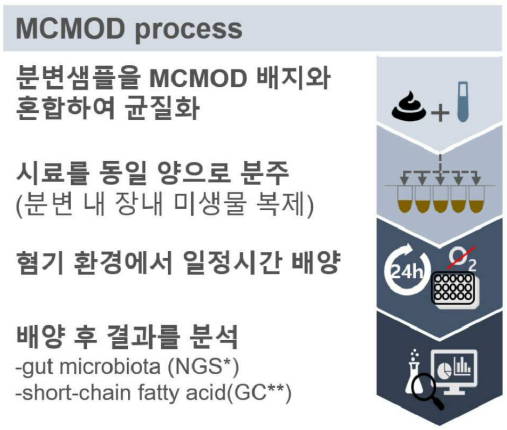
130: 진단부

도면

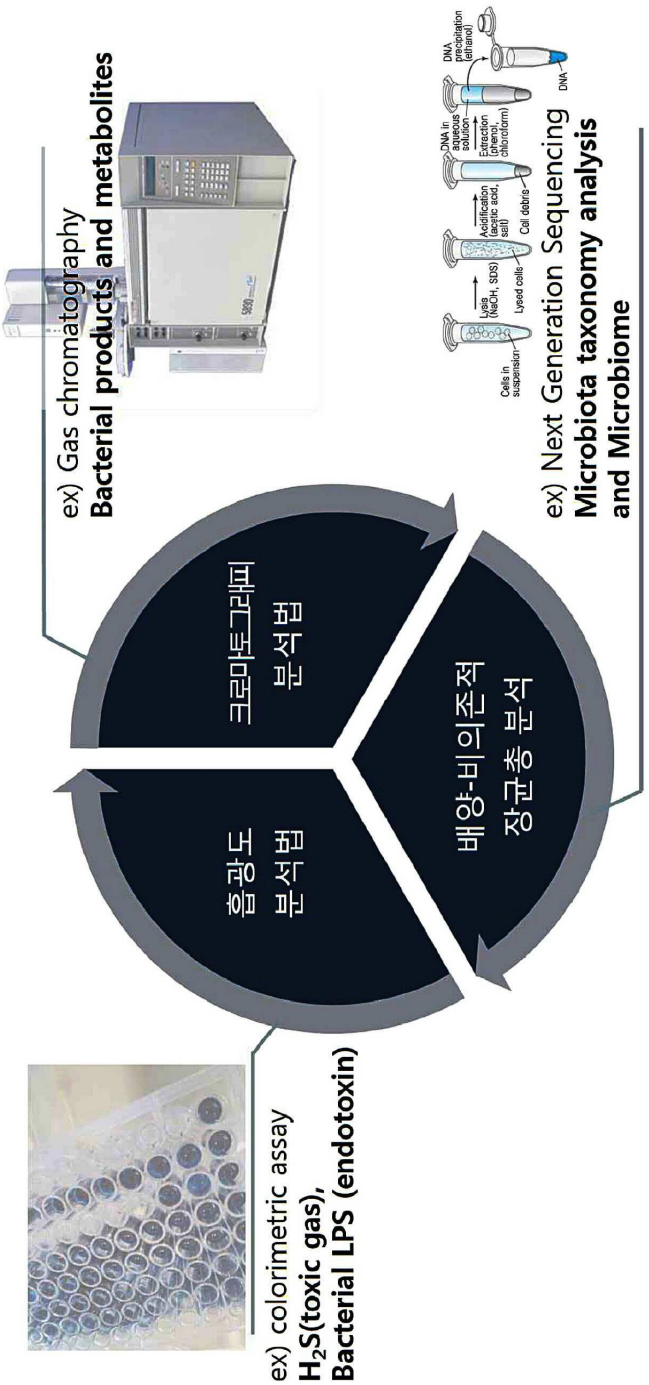
도면1



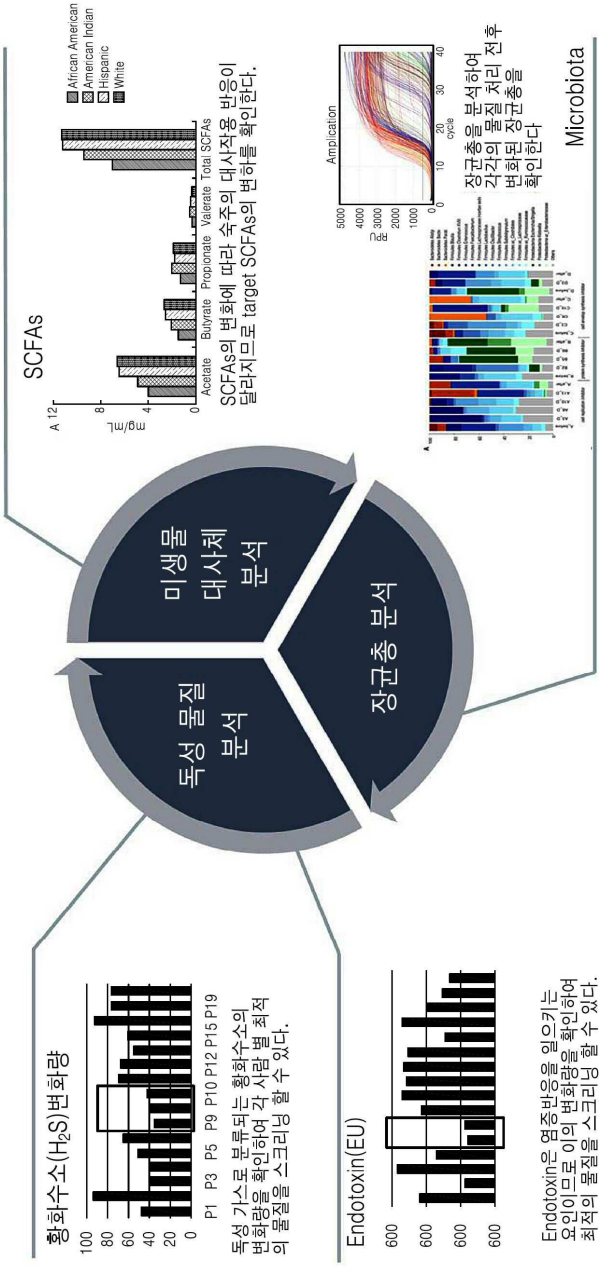
도면2


MCMOD process
분변샘플을 MCMOD 배지와 혼합하여 균질화
시료를 동일 양으로 분주
(분변 내 장내 미생물 복제)
혐기 환경에서 일정시간 배양
배양 후 결과를 분석
-gut microbiota (NGS*)
-short-chain fatty acid(GC**)
24h
O2

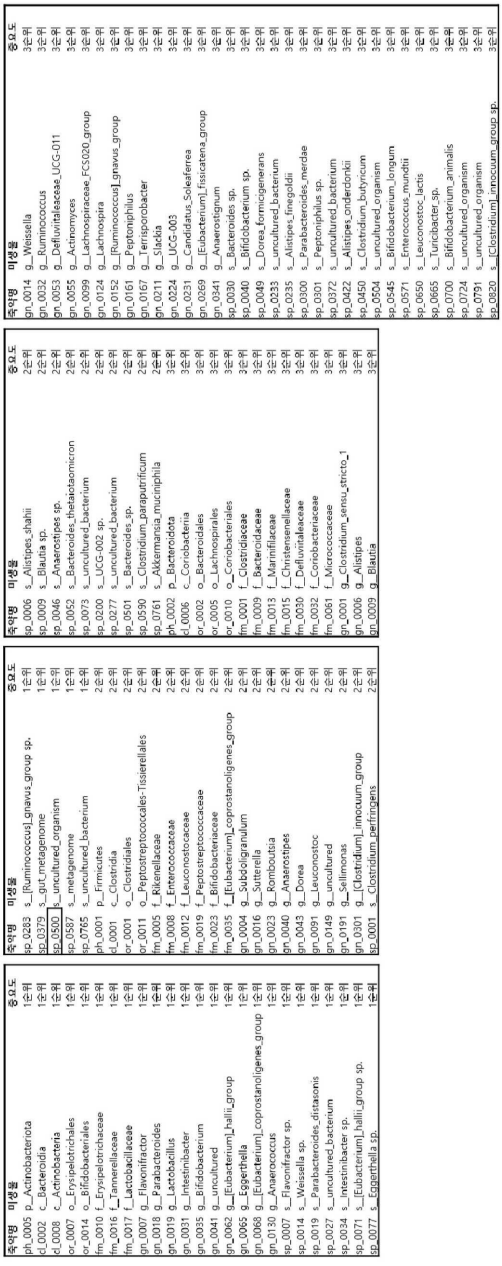
도면3


ex) Gas chromatography
Bacterial products and metabolites
크로마토그래피
분석법
흡광도
분석법
배양-비의존적
장균총 분석
ex) Next Generation Sequencing
Microbiota taxonomy analysis
and Microbiome
Cells in suspension
Lysis (NaOH, SDS)
Lysed cells
Acidification (acetic acid, salt)
Cell debris
DNA in aqueous solution
Extraction (pheno', chloroform)
DNA precipitation (ethanol)
DNA
ex) colorimetric assay
H2S(toxic gas),
Bacterial LPS (endotoxin)

도면4


SCFAs
African American
American Indian
Hispanic
White
Acetate
Butyrate
Propionate
Valerate
Total SCFAs
SCFAs의 변화에 따라 숙주의 대사작용 반응이 달라지므로 target SCFAs의 변화를 확인한다.
Amplication
cycle
RFU
장균총을 분석하여 각각의 물질처리 전후 변화된 장균총을 확인한다
Microbiota
미생물 대사체 분석
장균총 분석
독성 물질 분석
황화수소(H2S)변화량
P1 P3 P5 P9 P10 P12 P15 P19
독성 가스로 분류되는 황화수소의 변화량을 확인하여 각 사람 별로 최적의 물질을 스크리닝 할 수 있다.
Endotoxin(EU)
Endotoxin은 염증반응을 일으키는 요인이므로 이의 변화량을 확인하여 최적의 물질을 스크리닝 할 수 있다.

도면5



속약명	미생물	중요도
ph_0005	p__Actinobacteriota	1순위
cl_0002	c__Bacteroidia	1순위
cl_0008	c__Actinobacteria	1순위
or_0007	o__Erysipelotrichales	1순위
or_0014	o__Bifidobacteriales	1순위
fm_0010	f__Erysipelotrichaceae	1순위
fm_0016	f__Tannerellaceae	1순위
fm_0017	f__Lactobacillaceae	1순위
gn_0007	g__Flavonifractor	1순위
gn_0018	g__Parabacteroides	1순위
gn_0019	g__Lactobacillus	1순위
gn_0031	g__Intestinibacter	1순위
gn_0035	g__Bifidobacterium	1순위
gn_0041	g__uncultured	1순위
gn_0062	g__[Eubacterium]_hallii_group	1순위
gn_0065	g__Eggerthella	1순위
gn_0068	g__[Eubacterium]_coprostanoligenes_group	1순위
gn_0130	g__Anaerococcus	1순위
sp_0007	s__Flavonifractor sp.	1순위
sp_0014	s__Weissella sp.	1순위
sp_0019	s__Parabacteroides_distasonis	1순위
sp_0027	s__uncultured_bacterium	1순위
sp_0034	s__Intestinibacter sp.	1순위
sp_0071	s__[Eubacterium]_hallii_group sp.	1순위
sp_0077	s__Eggerthella sp.	1순위

속약명	미생물	중요도
sp_0283	s__[Ruminococcus]_gnavus_group sp.	1순위
sp_0379	s__gut_metagenome	1순위
sp_0500	s__uncultured_organism	1순위
sp_0587	s__metagenome	1순위
sp_0765	s__uncultured_bacterium	1순위
ph_0001	p__Firmicutes	2순위
cl_0001	c__Clostridia	2순위
or_0001	o__Clostridiales	2순위
or_0011	o__Peptostreptococcales-Tissierellales	2순위
fm_0005	f__Rikenellaceae	2순위
fm_0008	f__Enterococcaceae	2순위
fm_0012	f__Leuconostocaceae	2순위
fm_0019	f__Peptostreptococcaceae	2순위
fm_0023	f__Bifidobacteriaceae	2순위
fm_0035	f__[Eubacterium]_coprostanoligenes_group	2순위
gn_0004	g__Subdoligranulum	2순위
gn_0016	g__Sutterella	2순위
gn_0023	g__Romboutsia	2순위
gn_0040	g__Anaerostipes	2순위
gn_0043	g__Dorea	2순위
gn_0091	g__Leuconostoc	2순위
gn_0149	g__uncultured	2순위
gn_0191	g__Sellimonas	2순위
gn_0301	g__[Clostridium]_innocuum_group	2순위
sp_0001	s__Clostridium_perfringens	2순위

속약명	미생물	중요도
sp_0006	s__Alistipes_shahii	2순위
sp_0009	s__Blautia sp.	2순위
sp_0046	s__Anaerostipes sp.	2순위
sp_0052	s__Bacteroides_thetaiotaomicron	2순위
sp_0073	s__uncultured_bacterium	2순위
sp_0200	s__UCG-002 sp.	2순위
sp_0277	s__uncultured_bacterium	2순위
sp_0501	s__Bacteroides sp.	2순위
sp_0590	s__Clostridium_saccharolyticum	2순위
sp_0761	s__Akkermansia_muciniphila	2순위
ph_0002	p__Bacteroidota	3순위
cl_0006	c__Coriobacteriia	3순위
or_0002	o__Bacteroidales	3순위
or_0005	o__Lachnospirales	3순위
or_0010	o__Coriobacteriales	3순위
fm_0001	f__Clostridiaceae	3순위
fm_0009	f__Bacteroidaceae	3순위
fm_0013	f__Marinifilaceae	3순위
fm_0015	f__Christensenellaceae	3순위
fm_0030	f__Defluviitaleaceae	3순위
fm_0032	f__Coriobacteriaceae	3순위
fm_0061	f__Micrococcaceae	3순위
gn_0001	g__Clostridium_sensu_stricto_1	3순위
gn_0006	g__Alistipes	3순위
gn_0039	g__Blautia	3순위

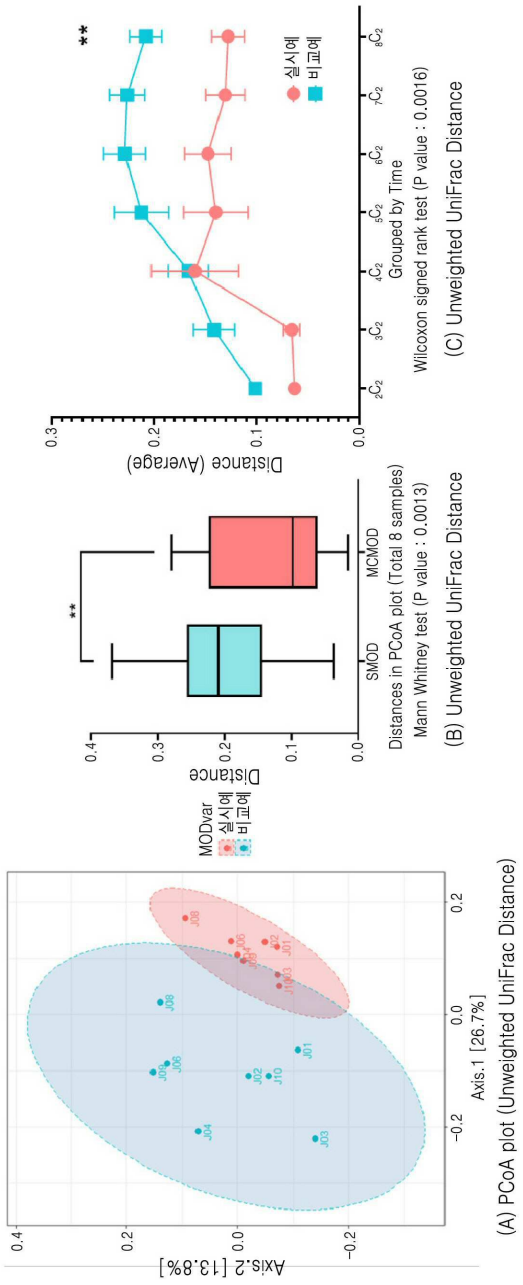
속약명	미생물	중요도
gn_0014	g__Weissella	3순위
gn_0032	g__Ruminococcus	3순위
gn_0053	g__Defluviitaleaceae_UCG-011	3순위
gn_0055	g__Actinomyces	3순위
gn_0099	g__Lachnospiraceae_FCS020_group	3순위
gn_0124	g__Lachnospira	3순위
gn_0152	g__[Ruminococcus]_gnavus_group	3순위
gn_0161	g__Peptoniphilus	3순위
gn_0167	g__Terrisporobacter	3순위
gn_0211	g__Slackia	3순위
gn_0224	g__UCG-003	3순위
gn_0231	g__Candidatus_Soleaferrea	3순위
gn_0269	g__[Eubacterium]_fissicatena_group	3순위
gn_0341	g__Anaerostipes	3순위
sp_0030	s__Bacteroides sp.	3순위
sp_0040	s__Bifidobacterium sp.	3순위
sp_0049	s__Dorea_formicigenerans	3순위
sp_0233	s__uncultured_bacterium	3순위
sp_0235	s__Alistipes_finegoldii	3순위
sp_0300	s__Parabacteroides_merdae	3순위
sp_0301	s__Peptoniphilus sp.	3순위
sp_0372	s__uncultured_bacterium	3순위
sp_0422	s__Alistipes_onderdonkii	3순위
sp_0450	s__Clostridium_butyricum	3순위
sp_0504	s__uncultured_organism	3순위
sp_0545	s__Bifidobacterium_longum	3순위
sp_0571	s__Enterococcus_mundtii	3순위
sp_0650	s__Leuconostoc_lactis	3순위
sp_0665	s__Turicibacter_sp.	3순위
sp_0700	s__Bifidobacterium_animalis	3순위
sp_0724	s__uncultured_organism	3순위
sp_0791	s__uncultured_organism	3순위
sp_0828	s__[Clostridium]_innocuum_group sp.	3순위

도면6

	Phylum	Class	Order	Family	Genus
ph_0001	Firmicutes				
ph_0002	Bacteroidota				
ph_0005	Actinobacteriota				
cl_0001	Firmicutes	Clostridia			
cl_0002	Bacteroidota	Bacteroidia			
cl_0006	Actinobacteriota	Coriobacteriia			
cl_0008	Actinobacteriota	Actinobacteria			
or_0001	Firmicutes	Clostridia	Clostridiales		
or_0002	Bacteroidota	Bacteroidia	Bacteroidales		
or_0005	Firmicutes	Clostridia	Lachnospirales		
or_0007	Firmicutes	Bacilli	Erysipelotrichales		
or_0010	Actinobacteriota	Coriobacteriia	Coriobacteriales		
or_0011	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales		
or_0014	Actinobacteriota	Actinobacteria	Bifidobacteriales		
fm_0001	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	
fm_0005	Bacteroidota	Bacteroidia	Bacteroidales	Rikenellaceae	
fm_0008	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	
fm_0009	Bacteroidota	Bacteroidia	Bacteroidales	Bacteroidaceae	
fm_0010	Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	
fm_0012	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	
fm_0013	Bacteroidota	Bacteroidia	Bacteroidales	Marinifilaceae	
fm_0015	Firmicutes	Clostridia	Christersenellales	Christensenellaceae	
fm_0016	Bacteroidota	Bacteroidia	Bacteroidales	Tannerellaceae	
fm_0017	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	
fm_0019	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	
fm_0023	Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	
fm_0030	Firmicutes	Clostridia	Lachnospirales	Defluviitaleaceae	
fm_0032	Actinobacteriota	Coriobacteriia	Coriobacteriales	Coriobacteriaceae	
fm_0035	Firmicutes	Clostridia	Oscillospirales	[Eubacterium]_coprostanoligenes_group	
fm_0061	Actinobacteriota	Actinobacteria	Micrococcales	Micrococcaceae	
gn_0001	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium_sensu_stricto_1
gn_0004	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	Subdoligranulum
gn_0006	Bacteroidota	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes
gn_0007	Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	Flavonifractor
gn_0009	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Blautia
gn_0014	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Weissella
gn_0016	Proteobacteria	Gammaproteobacteria	Burkholderiales	Sutterellaceae	Sutterella
gn_0018	Bacteroidota	Bacteroidia	Bacteroidales	Tannerellaceae	Parabacteroides
gn_0019	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
gn_0023	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	Romboutsia
gn_0031	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	Intestinibacter
gn_0032	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	Ruminococcus
gn_0035	Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium
gn_0040	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Anaerostipes
gn_0041	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	uncultured
gn_0043	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Dorea
gn_0053	Firmicutes	Clostridia	Lachnospirales	Defluviitaleaceae	Defluviitaleaceae_UCG-011
gn_0055	Actinobacteriota	Actinobacteria	Actinomycetales	Actinomycetaceae	Actinomyces
gn_0062	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	[Eubacterium]_halli_group
gn_0065	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Eggerthella
gn_0068	Firmicutes	Clostridia	Oscillospirales	[Eubacterium]_coprostanoligenes_group	[Eubacterium]_coprostanoligenes_group

	Phylum	Class	Order	Family	Genus	Species
gn_0091	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Leuconostoc	
gn_0099	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Lachnospiraceae_FCS020_group	
gn_0124	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Lachnospira	
gn_0130	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcales-Tissierellales	Anaerococcus	
gn_0149	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	uncultured	
gn_0152	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	[Ruminococcus]_gnavus_group	
gn_0161	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcales-Tissierellales	Peptoniphilus	
gn_0167	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	Terrisporobacter	
gn_0191	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Sellimonas	
gn_0211	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Slackia	
gn_0224	Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	UCG-003	
gn_0231	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	Candidatus_Soleaferrea	
gn_0269	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	[Eubacterium]_fissicatena_group	
gn_0301	Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	[Clostridium]_innocuum_group	
gn_0341	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Anaerostignum	
sp_0001	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium_sensu_stricto_1	Clostridium_perfringens
sp_0006	Bacteroidota	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	Alistipes_shahii
sp_0007	Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	Flavonifractor	NA
sp_0009	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Blautia	NA
sp_0014	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Weissella	NA
sp_0019	Bacteroidota	Bacteroidia	Bacteroidales	Tannerellaceae	Parabacteroides	Parabacteroides_distasonis
sp_0027	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	Subdoligranulum	uncultured_bacterium
sp_0030	Bacteroidota	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	NA
sp_0034	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	Intestinibacter	NA
sp_0040	Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	NA
sp_0046	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Anaerostipes	NA
sp_0049	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Dorea	Dorea_formicigenerans
sp_0052	Bacteroidota	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	Bacteroides_thetaiotaomicron
sp_0071	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	[Eubacterium]_hallii_group	NA
sp_0073	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	uncultured	uncultured_bacterium
sp_0077	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Eggerthella	NA
sp_0200	Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	UCG-002	NA
sp_0233	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium_sensu_stricto_1	uncultured_bacterium
sp_0235	Bacteroidota	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	Alistipes_finegoldii
sp_0277	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	uncultured	uncultured_bacterium
sp_0283	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	[Ruminococcus]_gnavus_group	NA
sp_0300	Bacteroidota	Bacteroidia	Bacteroidales	Tannerellaceae	Parabacteroides	Parabacteroides_merdae
sp_0301	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcales-Tissierellales	Peptoniphilus	NA
sp_0372	Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	Flavonifractor	uncultured_bacterium
sp_0379	Firmicutes	Clostridia	Oscillospirales	[Eubacterium]_coprostanoligenes_group	[Eubacterium]_coprostanoligenes_group	gut_metagenome
sp_0422	Bacteroidota	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	Alistipes_onderdonkii
sp_0450	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium_sensu_stricto_1	Clostridium_butyricum
sp_0500	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcales-Tissierellales	Peptoniphilus	uncultured_organism
sp_0501	Bacteroidota	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	Bacteroides_sp.
sp_0504	Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	UCG-003	uncultured_organism
sp_0545	Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	Bifidobacterium_longum
sp_0571	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus	Enterococcus_mundtii
sp_0587	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Blautia	metagenome
sp_0590	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium_sensu_stricto_1	Clostridium_paraputrificum
sp_0650	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Leuconostoc	Leuconostoc_lactis
sp_0665	Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	Turicibacter	Turicibacter_sp.
sp_0700	Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	Bifidobacterium_animalis
sp_0724	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Fusicatenibacter	uncultured_organism
sp_0761	Verrucomicrobiota	Verrucomicrobiae	Verrucomicrobiales	Akkermansiaceae	Akkermansia	Akkermansia_muciniphila
sp_0765	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Sellimonas	uncultured_bacterium
sp_0791	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	Ruminococcus	uncultured_organism
sp_0820	Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	[Clostridium]_innocuum_group	NA

도면7

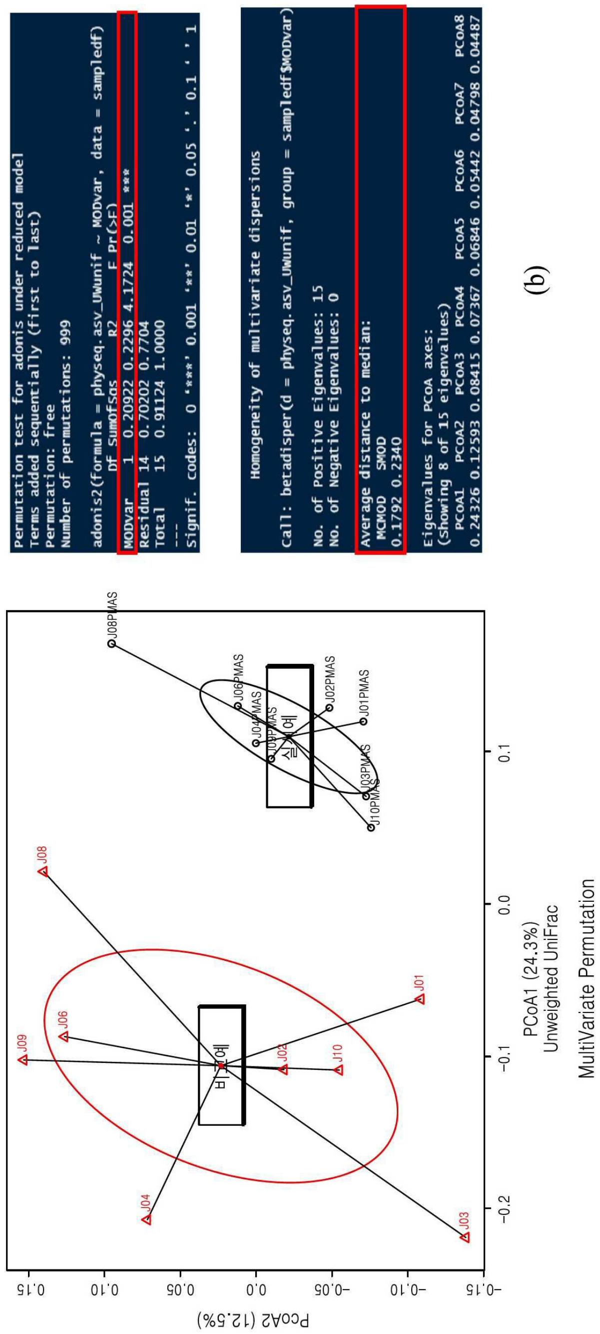


(A) PCoA plot (Unweighted UniFrac Distance)

(B) Unweighted UniFrac Distance

(C) Unweighted UniFrac Distance

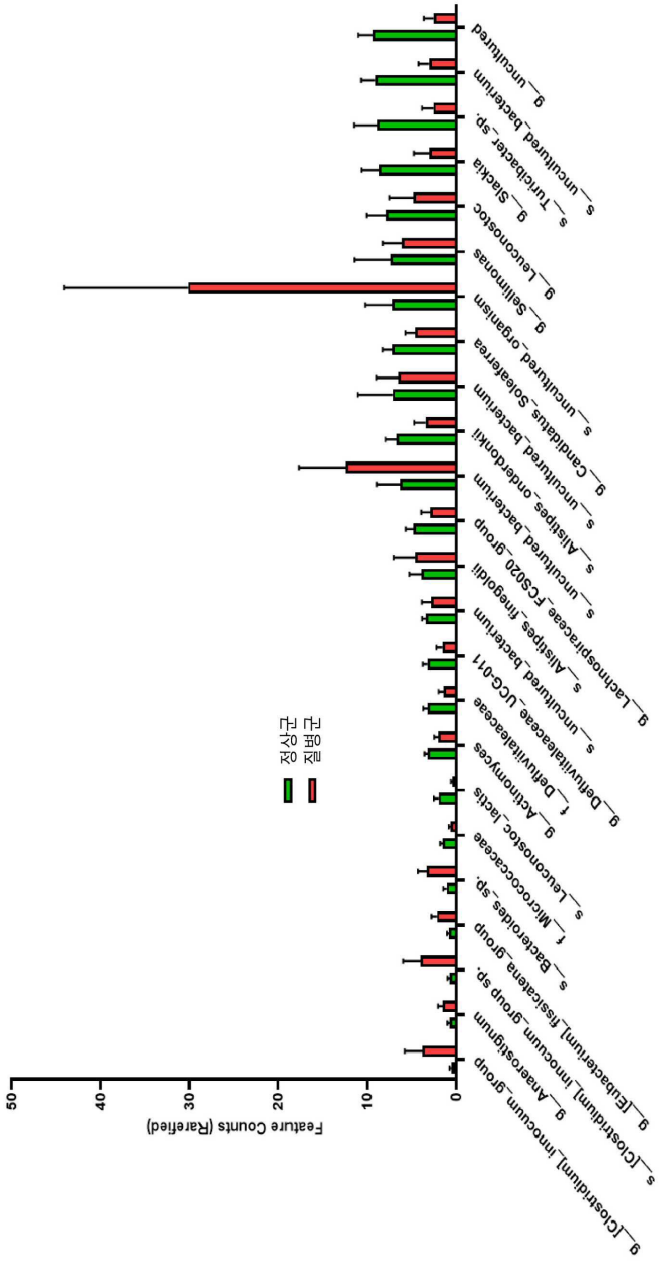
도면8


PcoA2 (12.5%)
0.15
0.10
0.05
0.0
−0.05
−0.10
−0.15
−0.2
−0.1
0.0
0.1
J09
J08
J06
J04
비교예
J02
J10
J01
J03
J08PMAS
J06PMAS
J04PMAS
J09PMAS
실시예
J02PMAS
J01PMAS
J03PMAS
J10PMAS
PCoA1 (24.3%)
Unweighted UniFrac
MultiVariate Permutation
(PERMANOVA)
Permutation test for adonis under reduced model
Terms added sequentially (first to last)
Permutation: free
Number of permutations: 999
adonis2(formula = physeq.asv_UWunif ~ MODvar, data = sampledf)
Df SumOfSqs R2 F Pr(>F)
MODvar 1 0.20922 0.2296 4.1724 0.001 ***
Residual 14 0.70202 0.7704
Total 15 0.91124 1.0000

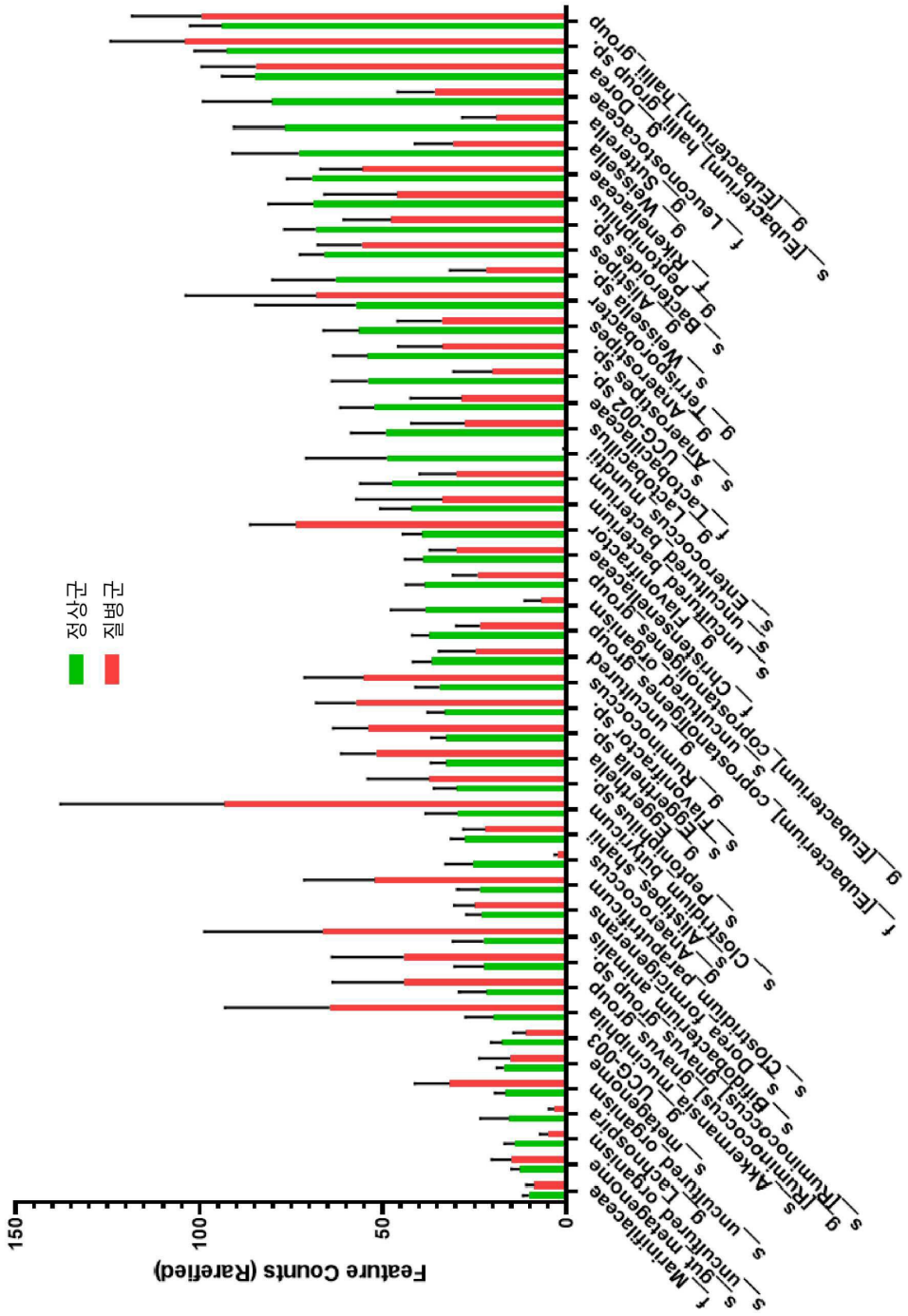
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Homogeneity of multivariate dispersions
Call: betadisper(d = physeq.asv_UWunif, group = sampledf$MODvar)
No. of Positive Eigenvalues: 15
No. of Negative Eigenvalues: 0
Average distance to median:
MCMOD SMOD
0.1792 0.2340
Eigenvalues for PCoA axes:
(Showing 8 of 15 eigenvalues)
PCoA1 PCoA2 PCoA3 PCoA4 PCoA5 PCoA6 PCoA7 PCoA8
0.24326 0.12593 0.08415 0.07367 0.06846 0.05442 0.04798 0.04487

(a) (b)

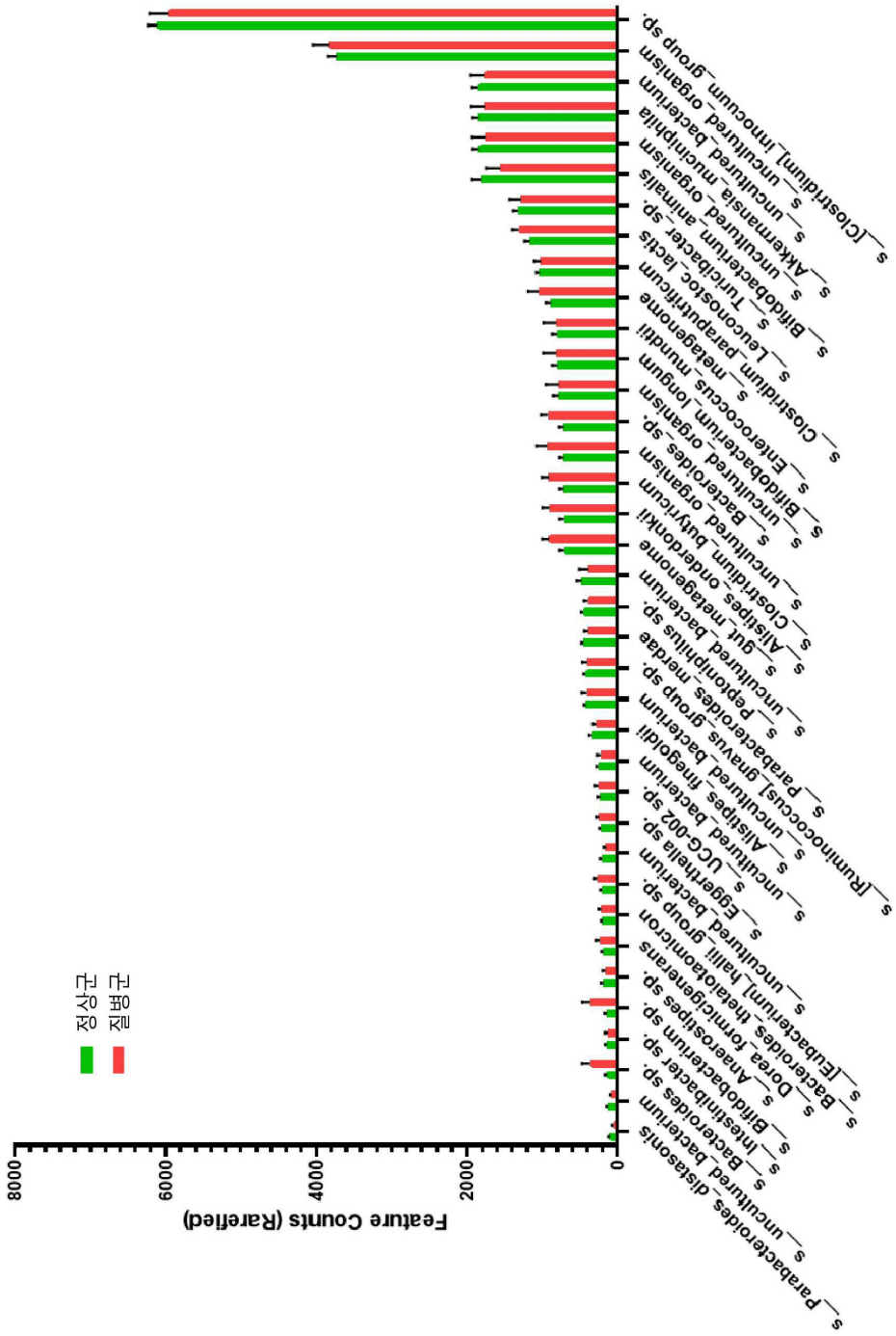
도면9


정상군
질병군
Feature Counts (Rarefied)
0
10
20
30
40
50
g__[Clostridium]_innocuum_group
g__Anaerostipes_hadrus
g__[Clostridium]_innocuum_group sp.
g__[Eubacterium]_fissicatena_group
s__Clostridium_sp.
g__Bacteroides
f__Micrococcaceae_sp.
g__Leuconostoc
s__Leuconostoc_lactis
g__Actinomyces
f__Defluviitaleaceae_UCG-011
g__Defluviitaleaceae_UCG-011
s__uncultured_bacterium
s__Alistipes_finegoldii
g__Lachnospiraceae_FCS020_group
s__uncultured_bacterium
s__Alistipes_onderdonkii
s__uncultured_bacterium
g__Candidatus_Soleaferrea
s__uncultured_organism
g__Sellimonas
g__Leuconostoc
g__Blautia
s__Turicibacter_sp.
s__uncultured_bacterium
g__uncultured

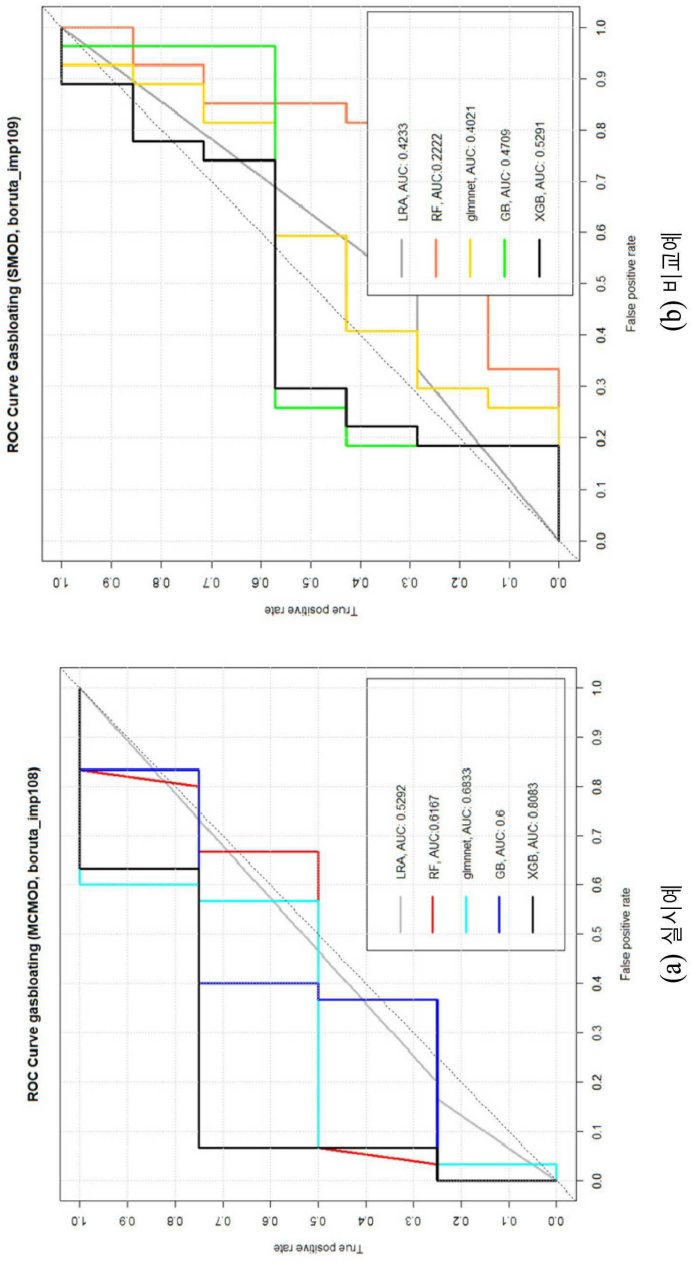
도면10


정상군
질병군
Feature Counts (Rarefied)
0
50
100
150
f_Marinifilaceae
s_gut_metagenome
s_uncultured_organism
g_Lachnospira
g_metagenome
g_UCG-003
s_uncultured_organism
g_Akkermansia
s_muciniphila
g_Ruminococcus_gnavus_group
s_Ruminococcus_gnavus
g_Blautia
s_Clostridium_formicigenerans
g_Dorea
s_Clostridium_paraputrificum
s_animalis
g_Anaerotruncus
s_Clostridium_sp.
g_Alistipes
s_shahii
g_Eggerthella
s_sp.
g_Faecalibacterium
s_prausnitzii
g_Ruminococcus_gnavus_group
s_sp.
g_uncultured_organism
g_uncultured_bacterium
g_Christensenellaceae_R-7_group
s_uncultured_bacterium
g_Coprostanoligenes_group
s_uncultured_bacterium
g_[Eubacterium]_coprostanoligenes_group
s_uncultured_bacterium
f_[Eubacterium]_coprostanoligenes_group
s_uncultured_bacterium
g_Enterococcus
s_mundtii
f_Lactobacillaceae
g_Lactobacillus
s_uncultured_bacterium
g_UCG-002
f_Ruminococcaceae
g_Anaerostipes
s_sp.
g_Terrisporobacter
s_Weissella_sp.
g_Bacteroides
s_sp.
g_Peptoniphilus
s_sp.
f_Rikenellaceae
g_Weissella
s_Sutterella
g_Leuconostocaceae
g_Dorea
s_sp.
f_[Eubacterium]_hallii_group
g_[Eubacterium]_hallii_group
s_sp.

도면11


정상군
질병군
Feature Counts (Rarefied)
0
2000
4000
6000
8000
s__Parabacteroides_distasonis
s__uncultured_Bacteroides
s__Intestinibacter_sp.
s__Bifidobacter_sp.
s__Dorea_formicigenerans
s__Anaerostipes_sp.
s__Bacteroides_thetaiotaomicron
s__[Eubacterium]_hallii_group_sp.
s__Eggerthella_sp.
s__uncultured_bacterium
s__UCG-002_sp.
s__uncultured_Alistipes_finegoldii
s__uncultured_gnavus_group_sp.
s__[Ruminococcus]_gnavus_group_sp.
s__Parabacteroides_merdae
s__Peptoniphilus_sp.
s__gut_metagenome
s__Alistipes_onderdonkii
s__Clostridium_butyricum
s__uncultured_Bacteroides_sp.
s__uncultured_Enterococcus_longum
s__Bifidobacterium_organism
s__Clostridium_metagenome
s__Enterococcus_faecium
s__uncultured_paraputrificum
s__Leuconostoc_lactis
s__Turicibacter_sp.
s__Bifidobacterium_animalis
s__Akkermansia_organism
s__uncultured_muciniphila
s__uncultured_bacterium
s__uncultured_organism
s__[Clostridium]_innocuum_group_sp.

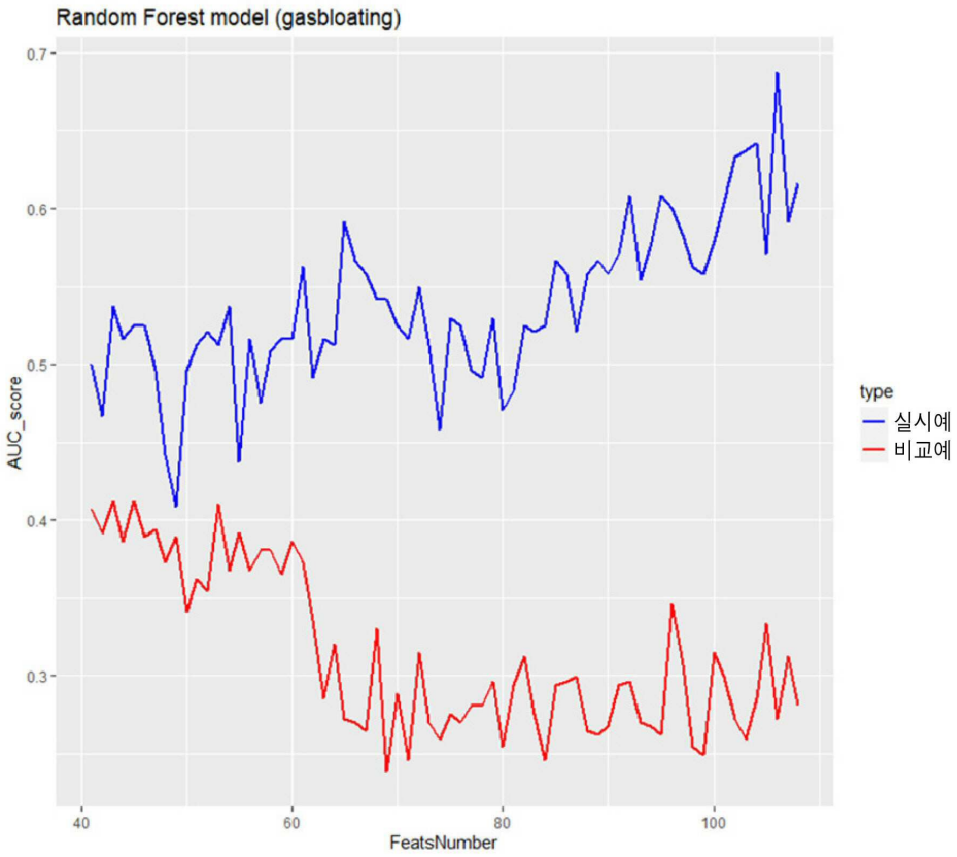
도면12



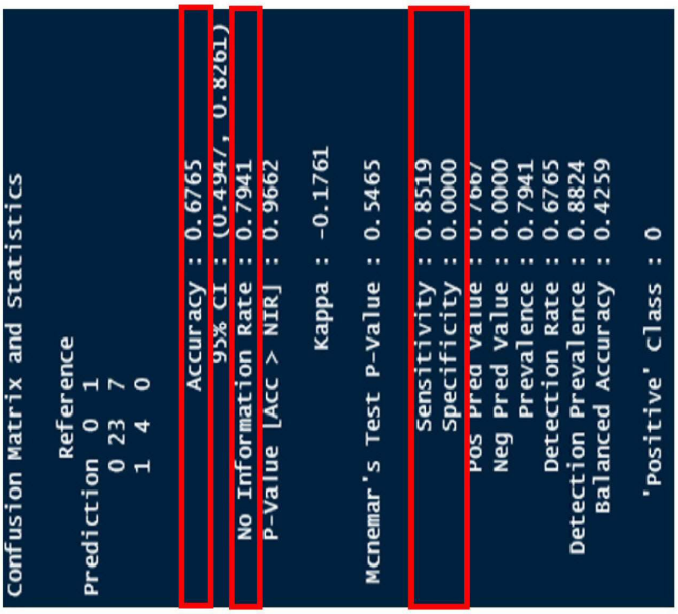
(b) 비교예

(a) 실시예

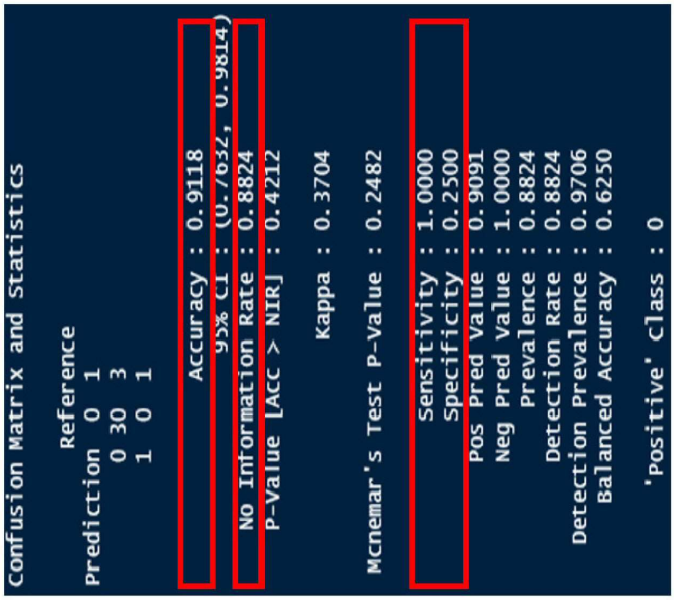
도면13


Random Forest model (gasbloating)
AUC_score
0.7
0.6
0.5
0.4
0.3
40
60
80
100
FeatsNumber
type
실시예
비교예

도면14

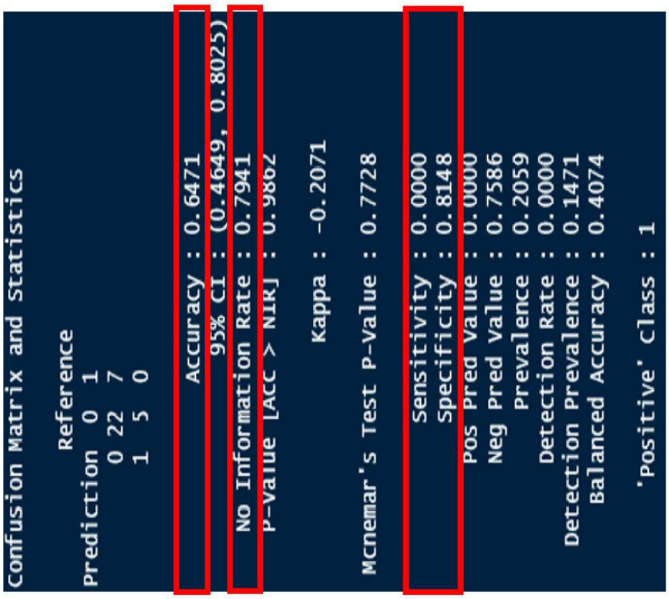


(b) 비교예

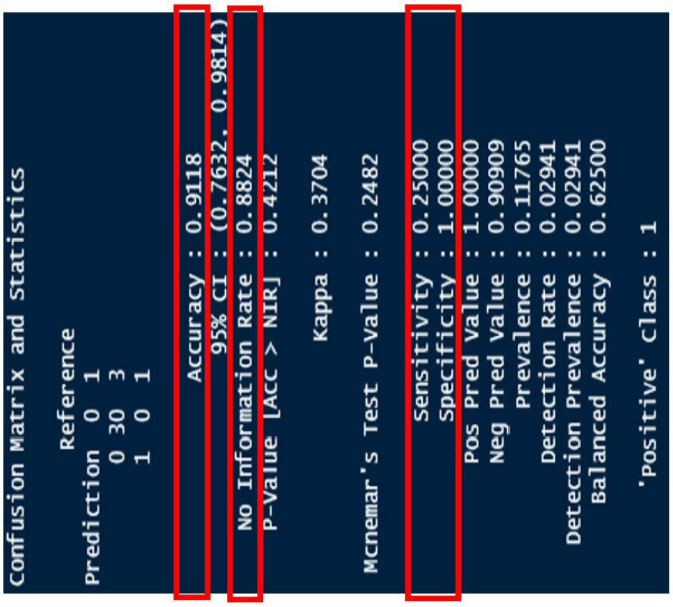


(a) 실시예

도면15

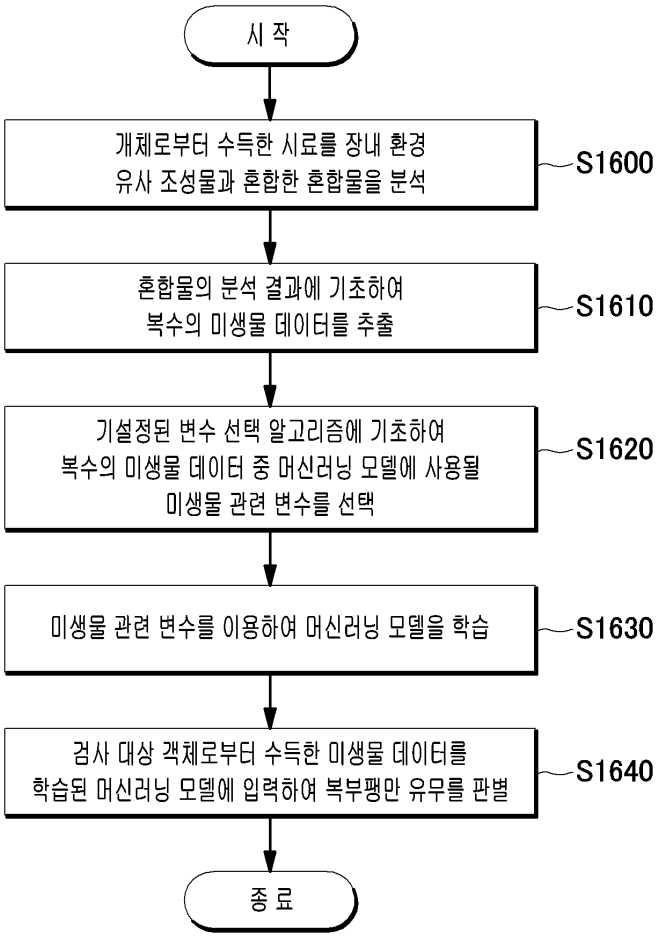


(b) 비교예



(a) 실시예

도면16


시 작
개체로부터 수득한 시료를 장내 환경 유사 조성물과 혼합한 혼합물을 분석
S1600
혼합물의 분석 결과에 기초하여 복수의 미생물 데이터를 추출
S1610
기설정된 변수 선택 알고리즘에 기초하여 복수의 미생물 데이터 중 머신러닝 모델에 사용될 미생물 관련 변수를 선택
S1620
미생물 관련 변수를 이용하여 머신러닝 모델을 학습
S1630
검사 대상 객체로부터 수득한 미생물 데이터를 학습된 머신러닝 모델에 입력하여 복부팽만 유무를 판별
S1640
종 료