

2021. 4. 20.(화) / 즉시

문의 : 담당자 연락처(02-880-4822)
연구단장/연구책임자 김희발 교수(02-880-4803) / 교신저자

서울대 김희발 교수팀, 장내 마이크로바이옴 기반 다중 질병 분류 기계학습 알고리즘을 개발하여 패컬티 오피니언 우수논문 선정

서울대 농업생명과학대학 김희발 교수팀과 (주)이지놈 (대표 조서애)은 2019년 네이처 자매지인 Scientific Reports에 발표된 논문이 2021년 패컬티 오피니언에 선정되었다고 밝혔다. 장내 마이크로바이옴은 숙주의 면역 시스템과 상호작용을 하여 숙주의 건강상태에 따라 달라질 수 있다. 본 연구는 정상인과 6가지 질병 (다발성 경화증, 소아 특발성 관절염, 만성피로 증후군, 에이즈, 뇌졸중, 직장암)을 가진 1,079명의 장내 마이크로바이옴의 구성을 분석하였고 4가지 다른 인공지능 분류 알고리즘으로 접근하여 관련 질병들을 구분해낼 수 있는 최적의 미생물종 및 분류법을 제시하였다 (그림 1). 이는 질병 예측 분야에서의 마이크로바이옴이 가질 수 있는 무궁무진한 가능성을 보여주는 사례이기도 하다.

본 논문은 인간유전체학 분야의 석학 스탠퍼드 의대 엘지자베스 멜린 교수의 추천하에 패컬티 오피니언스에 선정되었고 한국을 빛낸 사람들 (한빛사)에 소개되고 있다. 멜린 교수는 추천서를 통해 “장내 마이크로바이옴의 불균형이 질병의 단순한 생물지표인지 혹은 발병의 원인인지는 명확히 밝히지는 못했지만 인공지능 분야의 기계학습을 통하여 질병이 있는 환자의 마이크로바이옴이 확실히 구분됨을 보인 것이 논문의 우수성” 이라고 소개하고 있다.

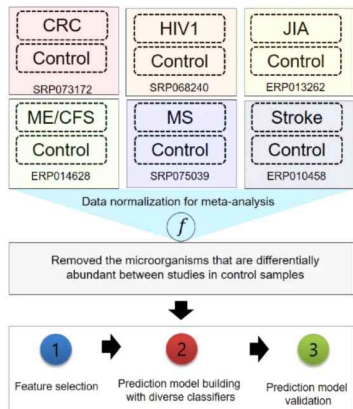


그림 1 여섯가지 질병에 대한 기계학습 기반 분류 모형 구축

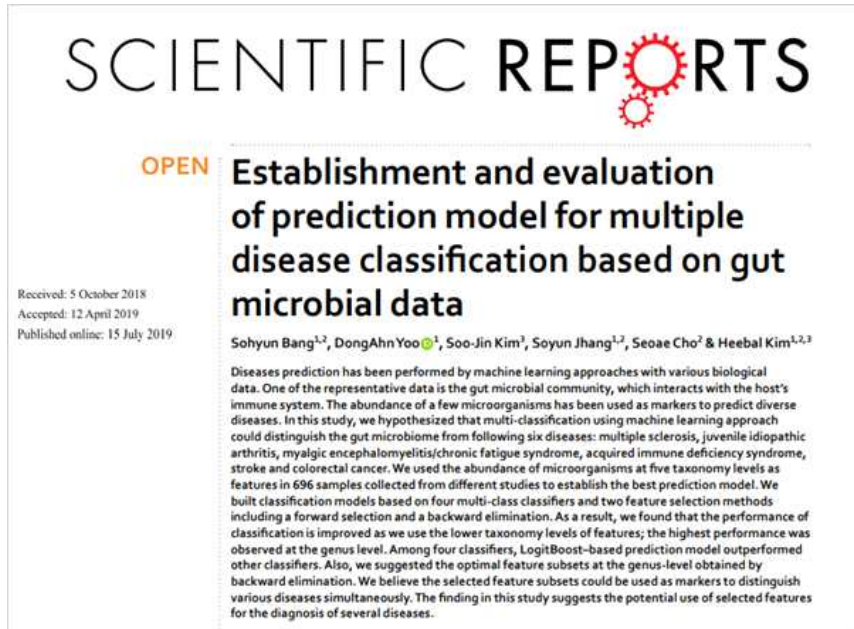
(주)이지놈의 조서애 대표는 “장내 마이크로바이옴이 인간의 질병 80% 이상 관여하고 있

어 질병별로 마이크로바이옴의 특이한 불균형을 프로파일링하는 것은 질병 치료와 생물지표를 발굴하는데 직접적인 정보를 제공하고 있다” 고 설명하였다.

본 연구는 식품의약품안전처의 지원에 의해 (과제번호: 14162MFDS972) 이루어진 것이며 서울대학교 김희발 교수팀과 (주)이지놈의 공동 연구로 진행되었다.

[붙임] 1. 연구결과

연구 결과



(2019년 7월 15일, Scientific reports 게재)

Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data

근래 이르러 차세대 염기서열 분석 기술의 발달과 함께 대량의 마이크로바이옴 자료들이 생산 및 축적되어왔으며 본 연구는 축적된 마이크로바이옴 자료들을 모아 메타 분석법을 통해 자료들의 재해석을 시도하였다. 공공 데이터베이스에 공개되어 있는 자료들로 다발성 경화증, 소아 특발성 관절염, 만성피로 증후군, 에이즈, 뇌졸중, 직장암을 포함한 여섯가지 질병 샘플들이 있으며, 이러한 다양한 질병은 장내 마이크로바이옴 커뮤니티에도 직/간접적인 영향을 끼칠 수 있다고 알려져 있다. 본 연구는 이러한 질병이 가져오는 장내 마이크로바이옴 커뮤

니티의 비대칭을 기계학습적 방법을 적용하여 분석한 연구이며, 여러 질병을 구분하기 위해 마이크로바이옴 자료가 사용된 몇 안되는 연구 사례이다.

본 연구는 상기 언급된 6가지 질병을 가진 환자들의 장내 마이크로바이옴을 문(Phylum) 에서 속 (Genus) 수준에 이르기까지 다양한 피쳐셋 (feature set)을 대상으로 LogitBoost, LMT (Logistic model tree), SVM (Support vector machine), KNN (K-nearest neighbor)를 포함한 기계학습 방법으로 분류하였으며, 장내 마이크로바이옴만을 이용하여 이러한 여러 질병을 높은 정확도로 구분할 수 있음을 보였다 (표1).

	CRC	HIV1	JIA	ME/CFS	MS	Stroke	Average
Accuracy							
LogitBoost	96.84 ± 0.43	99.71 ± 0.14	98.52 ± 0.22	96.93 ± 0.46	98.28 ± 0.29	98.32 ± 0.46	98.1 ± 0.33
LMT	95.93 ± 0.3	98.66 ± 0.22	98.95 ± 0.22	96.26 ± 0.57	98.18 ± 0.44	98.8 ± 0.22	97.8 ± 0.33
SVM	95.59 ± 0.5	98.85 ± 0.25	98.28 ± 0.38	96.46 ± 0.08	98.08 ± 0.22	98.75 ± 0.22	97.67 ± 0.28
KNN	90.28 ± 0.3	97.27 ± 0.43	97.27 ± 0	94.73 ± 0.36	96.41 ± 0.14	96.55 ± 0.5	95.42 ± 0.29
FPR							
	CRC	HIV1	JIA	ME/CFS	MS	Stroke	Average
LogitBoost	3.7 ± 0.83	0.26 ± 0.11	0.85 ± 0.09	1.18 ± 0.24	0.4 ± 0.09	1.14 ± 0.28	1.26 ± 0.27
LMT	3.93 ± 0.4	0.85 ± 0.3	0.6 ± 0	1.7 ± 0.41	0.7 ± 0.43	0.9 ± 0.18	1.45 ± 0.29
SVM	4.77 ± 0.48	0.59 ± 0.2	0.8 ± 0.09	1.59 ± 0.09	0.9 ± 0.15	0.6 ± 0.1	1.54 ± 0.19
KNN	12.93 ± 0.23	1.83 ± 0.49	1.35 ± 0.15	0.87 ± 0.32	0.4 ± 0.17	2.34 ± 0.18	3.29 ± 0.26
FNR							
	CRC	HIV1	JIA	ME/CFS	MS	Stroke	Average
LogitBoost	2.28 ± 0.38	0.36 ± 0.31	16.09 ± 3.98	28.47 ± 9.62	32.18 ± 7.18	3.78 ± 1.48	13.86 ± 3.82
LMT	4.31 ± 0.22	2.69 ± 0	11.49 ± 5.27	31.25 ± 3.61	27.59 ± 3.45	2.36 ± 1.64	13.28 ± 2.37
SVM	3.8 ± 0.66	2.69 ± 1.08	22.99 ± 7.18	29.86 ± 1.2	25.29 ± 1.99	3.78 ± 0.82	14.74 ± 2.16
KNN	4.44 ± 0.44	5.2 ± 0.31	34.48 ± 3.45	64.58 ± 2.08	77.01 ± 5.27	7.8 ± 2.56	32.25 ± 2.35

표 1 여섯가지 질병에 대한 장내 마이크로바이옴 기계학습 분류법의 정확도 및 1/2종 오류

본 연구는 장내 마이크로바이옴의 불균형이 다수의 질병에서 서로 다른 양상을 보이며, 몇몇 미생물들의 경우 다수의 질병 예측 목적의 지표로서 높은 잠재성을 가질 수 있음을 제시하고 있다.