



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년10월29일
(11) 등록번호 10-2723638
(24) 등록일자 2024년10월25일

- (51) 국제특허분류(Int. Cl.)
G16B 40/20 (2019.01) G06N 20/00 (2019.01)
G16B 30/00 (2019.01)
- (52) CPC특허분류
G16B 40/20 (2019.02)
G06N 20/00 (2021.08)
- (21) 출원번호 10-2024-0004071
(22) 출원일자 2024년01월10일
심사청구일자 2024년01월10일
- (56) 선행기술조사문헌
Edward J. Hu 외, Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685. (2021. 10. 16.)*
Sung-Yoon Ahn 외, Toxic Fungi Protein Classification Using Task Specific BERT, The 8th International Conference on Next Generation Computing 2022, pp.319-320. (2022. 10.)*
*는 심사관에 의하여 인용된 문헌

- (73) 특허권자
조선대학교 산학협력단
광주광역시 동구 조선대길 126(서석동, 산학협력단)
- (72) 발명자
방일수
광주광역시 동구 조선대길 7, 치과대학 연구동 4212호(서석동)
이상윤
서울특별시 강남구 삼성로 212, 7동 908호 (대치동, 은마아파트)
안성윤
경기도 성남시 수정구 복정로16번길 22, 304호(복정동)
- (74) 대리인
김견수

전체 청구항 수 : 총 4 항

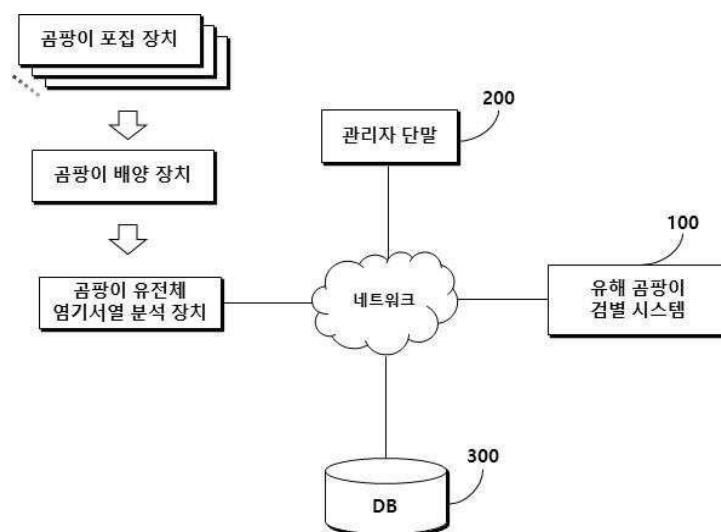
심사관 : 성경아

(54) 발명의 명칭 실내 유해 곰팡이 검별 시스템 및 그 방법

(57) 요약

본 발명은 실내 유해 곰팡이 검별 시스템에 관한 것으로, 사전 학습된 학습모델에서 특정 레이어의 가중치를 선택하여 파인튜닝하고, 상기 파인튜닝한 사전 학습된 학습모델을 유해 곰팡이 검별을 위한 곰팡이 유전체 염기서열 학습데이터를 이용하여 학습시켜 유해 곰팡이 검별용 학습모델을 생성함으로써 상기 유해 곰팡이 검별용 학습모델을 경량화하며, 실제 공기 중 포집한 곰팡이의 유전체 염기서열을 상기 유해 곰팡이 검별용 학습모델에 입력하여 유해 곰팡이를 신속하게 검별할 수 있도록 하는 시스템 및 그 방법에 관한 것이다.

대표도 - 도1



(52) CPC특허분류

G16B 30/00 (2019.02)

이 발명을 지원한 국가연구개발사업

과제고유번호 1485019431

과제번호 ARQ202101648003

부처명 환경부

과제관리(전문)기관명 한국환경산업기술원

연구사업명 실내공기생물학적위해인자관리기술개발사업(034122007007922002205314)

연구과제명 실내공기 생물 유해인자에 의한 세포 및 생물체 반응 평가를 위한 다각적 모델링 기

법 개발

과제수행기관명 조선대학교 산학협력단

연구기간 2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

실내 유해 곰팡이 검별 시스템에서, 사전에 수집한 곰팡이별 유전체 염기서열에 각 곰팡이가 독소를 생성하는지, 독성을 가지는지 혹은 인체에 무해한지에 대한 레이블을 각각 레이블링함으로써 곰팡이 유전체 염기서열 학습데이터를 생성하는 학습데이터 생성 단계;

상기 실내 유해 곰팡이 검별 시스템에서, 상기 생성한 학습데이터를 활용하여 사전 학습된 학습모델을 학습시킴으로써 유해 곰팡이 검별용 학습모델을 생성하는 유해 곰팡이 검별용 학습모델 생성 단계;

상기 실내 유해 곰팡이 검별 시스템에서, 곰팡이 유전체 염기서열 분석 장치로부터 실내 공기 중에서 포집된 곰팡이의 유전체 염기서열을 수신하는 곰팡이 유전체 염기서열 수신 단계; 및

상기 실내 유해 곰팡이 검별 시스템에서, 상기 수신한 곰팡이의 유전체 염기서열을 상기 유해 곰팡이 검별용 학습모델에 입력하여 상기 곰팡이가 유해 곰팡이인지를 검별하는 유해 곰팡이 검별 단계;를 포함하며,

상기 유해 곰팡이 검별용 학습모델은,

상기 곰팡이 유전체 염기서열로부터 곰팡이의 종류를 분류하기 위해 상기 사전 학습된 학습모델을 기반으로 생성되며,

상기 생성한 곰팡이 유전체 염기서열 학습데이터를 활용하여 상기 사전 학습된 학습모델을 학습함에 있어서, 상기 사전 학습된 학습모델에서 특정 레이어의 사전 학습된 가중치(W) 메트릭을 구성하는 가중치는 고정하며,

상기 특정 레이어의 사전 학습된 가중치 메트릭에서, 적어도 하나 이상의 소정의 가중치 메트릭을 선택하여 로우 랭크 디컴포지션(low rank decomposition)하고,

상기 사전 학습된 학습모델을 학습시키는 과정에서 누적 업데이트되는 그레디언트(gradient)에 따라 상기 로우 랭크 디컴포지션한 소정의 가중치 메트릭을 구성하는 가중치를 업데이트하여, 상기 로우 랭크 디컴포지션한 소정의 가중치 메트릭을 곱한 다음, 상기 곱한 결과를 상기 가중치를 고정한 상기 선택한 소정의 가중치 메트릭에 합산함으로써, 상기 사전 학습된 학습모델을 파인튜닝(fine tuning)하는 것을 포함하며,

상기 파인튜닝을 통해서 곰팡이 유전체 염기서열이 입력되면, 상기 사전 학습된 학습모델에서 곰팡이 유전체 염기서열을 학습할 때 이미 학습된 Q(query), K(key), V(value) 중 K(key)에 대해서는 다시 학습하지 않도록 하여, 상기 학습을 수행하는 학습 파라미터를 줄임으로써, 상기 유해 곰팡이 검별용 학습모델을 경량화하며,

상기 K(key)는, 상기 입력된 곰팡이 유전체 염기서열을 구성하는 각 핵 염기간의 관련성 및 유사성을 판단하는 기준이 되는 것을 특징으로 하는 실내 유해 곰팡이 검별 방법.

청구항 2

삭제

청구항 3

삭제

청구항 4

청구항 1에 있어서,

상기 유해 곰팡이 검별 단계는,

상기 유해 곰팡이 검별용 학습모델의 출력결과 중 확률이 제일 높은 레이블을 선택함으로써 유해 곰팡이를 검별하는 것을 포함하는 것을 특징으로 하는 실내 유해 곰팡이 검별 방법.

청구항 5

청구항 1에 있어서,

상기 실내 유해 곰팡이 검별 방법은,

상기 수신한 곰팡이 유전체 염기서열이 소정의 길이 미만이면, 상기 길이에 맞게 0을 패딩하여 상기 수신한 곰팡이 유전체 염기서열을 전처리하는 전처리 단계;를 더 포함하는 것을 특징으로 하는 실내 유해 곰팡이 검별 방법.

청구항 6

청구항 1, 청구항 4 또는 청구항 5에 따른 실내 유해 곰팡이 검별 방법을 수행하는 프로그램 코드를 저장하는 메모리; 및

상기 메모리에 저장된 상기 프로그램 코드를 실행하도록 구성된 프로세서;를 포함하는 것을 특징으로 하는 실내 유해 곰팡이 검별 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 실내 유해 곰팡이 검별 시스템에 관한 것으로, 더욱 상세하게는 사전 학습된 학습모델에서 특정 레이어의 가중치를 선택하여 파인튜닝하고, 상기 파인튜닝한 사전 학습된 학습모델을 유해 곰팡이 검별을 위한 곰팡이 유전체 염기서열 학습데이터를 이용하여 학습시켜 유해 곰팡이 검별용 학습모델을 생성함으로써 상기 유해 곰팡이 검별용 학습모델을 경량화하며, 실제 공기 중 포집한 곰팡이의 유전체 염기서열을 상기 유해 곰팡이 검별용 학습모델에 입력하여 유해 곰팡이를 신속하게 검별할 수 있도록 하는 시스템 및 그 방법에 관한 것이다.

배경 기술

- [0002] 기술의 발전으로 인해 인간의 생활 및 작업 환경은 실외에서 실내로 점차 이동되고 있다.
- [0003] 실내공간은 인간에게 기상과 같은 다양한 외부 자극으로부터 보호하여 안정적인 환경을 제공하지만, 습하고 따뜻한 실내공기는 곰팡이의 이상적인 번식환경을 제공하여 천식 등과 같은 다양한 건강문제를 유발할 수 있다.
- [0004] 예를 들어, 일반적인 가정에서 발생할 수 있는 곰팡이의 한 종류인 아스페르길루스(aspergillus)는 면역 체계가 약화된 사람들에게 알레르기 반응이나 폐 질환을 발생시킬 수 있으며, 또 다른 곰팡이의 한 종류인 주폐포자충(pneumocystis jirovecii)은 주폐포자충 폐렴이라는 심각한 질병을 유발할 수 있다.
- [0005] 따라서 실내공기에 대한 적절한 환기도 중요하지만, 광범위한 질병을 예방하기 위해서는 실내 공기 질에 대한 종합적인 점검이 필요하다.
- [0006] 일반적으로 실내 공기 질에 대한 점검은 실내 공기에서 포집한 곰팡이의 유전체 염기서열(단백질 서열)에 대한 시험관 내 분석(in-vitro)을 통해 유해 곰팡이를 찾는 방식으로 수행된다.
- [0007] 상기 유전체 염기서열의 시험관 내 분석은 정확한 결과를 제공하는 장점이 있으나 시간이 오래 걸리며 비용이 많이 드는 문제점이 있다.
- [0008] 이에 따라 본 발명에서는 곰팡이의 유전체 염기서열에 유해한 곰팡이임을 나타내는 레이블을 레이블링한 곰팡이 유전체 염기서열 학습데이터를 사전에 학습된 학습모델을 통해 학습하여 유해한 곰팡이를 신속 정확하게 검별할 수 있도록 하는 방안을 제안하고자 한다.
- [0009] 또한 본 발명은 사전에 학습된 학습모델의 특정 레이어의 가중치를 선택하여 학습시키도록 상기 사전 학습된 학습모델을 파인튜닝함으로써 유해 곰팡이를 신속하게 검별할 수 있도록 하는 곰팡이 검별용 학습모델을 생성하는 방안을 제안하고자 한다.
- [0010] 이때, 본 발명은 사전 학습된 학습모델의 기존 가중치는 고정된 채로, 특정 레이어의 사전 학습된 가중치 메트릭(metric)을 구성하는 가중치를 고정하고, 상기 사전 학습된 가중치 메트릭 중 소정의 가중치 메트릭을 선택하여 로우 랭크 디컴포지션(low rank decomposition)하고 학습시 로우 랭크 디컴포지션된 가중치 메트릭의 가중치만 학습시켜 기존 가중치에 합산하도록 상기 사전 학습된 학습모델을 파인튜닝함으로써 사전 학습된 학습모델보다 적은 용량으로 보다 정확한 유해 곰팡이를 검별하는 곰팡이 검별용 학습모델을 생성할 수 있도록 하는 방

안을 제안하고자 한다.

- [0011] 다음으로 본 발명의 기술분야에 존재하는 선행기술에 대하여 간단하게 설명하고, 이어서 본 발명이 상기 선행기술에 비해서 차별적으로 이루고자 하는 기술적 사항에 대해서 기술하고자 한다.
- [0012] 먼저 한국등록특허 제2005674호(2019.07.24.)는 공기 중 바이러스 및 곰팡이 측정 방법에 관한 것으로, 포집부를 통해 곰팡이를 포집하여 아가플레이트와 배양기를 통해 배양하고 배양한 곰팡이를 측정하는 공기 중 바이러스 및 곰팡이 측정 방법에 관한 것이다.
- [0013] 즉, 한국등록특허 제2005674호는 배양한 곰팡이의 개수를 확인함으로써 곰팡이를 측정하는 정도만 기재하고 있을 뿐, 본 발명에서 기재하고 있는 실내 공기의 곰팡이를 포집하여 유행 곰팡인지 여부를 검별하는 방법에 대해서는 전혀 기재하거나 시사 혹은 그 어떠한 암시도 없다.
- [0014] 또한 한국공개특허 제2023-0144967호(2023.10.17.)는 공기 중의 다중 곰팡이 검출을 위한 휴대용 검출 장치에 관한 것으로, 공기를 흡입하여 곰팡이를 포집하고, 포집한 곰팡이를 파쇄한 후, 파쇄한 곰팡이 포집 용액으로부터 곰팡이를 검출하여 농도를 분석하는 공기 중의 다중 곰팡이 검출을 위한 휴대용 검출 장치에 관한 것이다.
- [0015] 즉, 한국공개특허 제2023-0144967호는 공기 중 곰팡이를 검출하고 농도를 분석하는 정도만 기재하고 있을 뿐이다.
- [0016] 반면에 본 발명은 곰팡이 유전체 염기서열에 독소 생성 혹은 독성을 가지는지 여부에 따른 레이블을 부가하여 레이블링한 곰팡이 유전체 염기서열 학습데이터를 생성하고, 상기 생성한 곰팡이 유전체 염기서열 학습데이터를 사전 학습한 학습모델을 통해 학습하여 생성한 유해 곰팡이 검별용 학습모델을 생성하며, 실제 실내 공기 중에서 포집한 곰팡이의 유전체 염기서열을 상기 곰팡이 검별용 학습모델에 입력하여 유해 곰팡이를 검별하는 것으로, 한국공개특허 제2023-0144967호는 이러한 본 발명의 기술적 특징을 기재하거나 시사 혹은 그 어떠한 암시도 없다. 따라서 양 발명은 그 기술적 구성, 목적 및 효과에 있어서 현저한 차이점이 있다.

발명의 내용

해결하려는 과제

- [0017] 본 발명은 상기와 같은 문제점을 해결하기 위해 창작된 것으로서, 공기 중에 포집한 적어도 하나 이상의 곰팡이에 대한 각 유전체 염기서열을 유해 곰팡이 검별용 학습모델에 입력하여 유해 곰팡이를 신속하게 검별할 수 있도록 하는 실내 유해 곰팡이 검별 시스템 및 그 방법을 제공하는 것을 그 목적으로 한다.
- [0018] 또한 본 발명은 유해 곰팡이 검별용 학습모델을 생성하기 위해서 곰팡이별 유전체 염기서열에 상기 곰팡이가 독소를 생성하는지, 독소에 의한 독성을 가지는지, 혹은 무해한 것인지에 대한 레이블을 레이블링하여 곰팡이 유전체 염기서열 학습데이터를 생성하는 시스템 및 그 방법을 제공하는 것을 그 목적으로 한다.
- [0019] 또한 본 발명은 사전 학습된 학습모델을 파인튜닝하여 곰팡이 유전체 염기서열 학습데이터를 학습함으로써 유해 곰팡이 검별용 학습모델을 생성하는 시스템 및 그 방법을 제공하는 것을 그 목적으로 한다.
- [0020] 또한 본 발명은 사전 학습된 학습모델에서, 유해 곰팡이 검별을 위한 적어도 하나 이상의 특정 레이어에서 소정의 가중치 메트릭을 로우 랭크 디콤폰지션하고, 상기 학습과정에서 로우 랭크 디콤폰지션된 메트릭의 가중치만 학습시켜, 기존 가중치에 합산하도록 함으로써 사전 학습된 학습모델을 파인튜닝하는 시스템 및 그 방법을 제공하는 것을 그 목적으로 한다.
- [0021] 또한 본 발명은 유해 곰팡이 검별용 학습모델을 통해 시험관 내 분석보다 신속하게 유해 곰팡이를 검별함으로써 실내 공기를 신속하게 정화할 수 있도록 하는 시스템 및 그 방법을 제공하는 것을 그 목적으로 한다.

과제의 해결 수단

- [0022] 본 발명의 일 실시예에 따른 실내 유해 곰팡이 검별 시스템은, 곰팡이 유전체 염기서열 분석 장치로부터 실내 공기 중에서 포집된 곰팡이의 유전체 염기서열을 수신하는 곰팡이 유전체 염기서열 수신부 및 상기 수신한 곰팡이의 유전체 염기서열을 유해 곰팡이 검별용 학습모델에 입력하여 상기 곰팡이가 유해 곰팡이인지를 검별하는 유해 곰팡이 검별부를 포함하며, 상기 유해 곰팡이 검별용 학습모델은, 상기 곰팡이 유전체 염기서열로부터 곰팡이의 종류를 분류하기 위해 사전 학습된 학습모델을 기반으로 생성되는 것을 특징으로 한다.
- [0023] 또한 상기 실내 유해 곰팡이 검별 시스템은, 사전에 수집한 곰팡이별 유전체 염기서열에 각 곰팡이가 독소를 생

성하는지, 독성을 가지는지 혹은 인체에 무해한지에 대한 레이블을 각각 레이블링함으로써 곰팡이 유전체 염기서열 학습데이터를 생성하는 학습데이터 생성부 및 상기 생성한 학습데이터를 활용하여 상기 사전 학습된 학습모델을 학습시킴으로써 상기 유해 곰팡이 검별용 학습모델을 생성하는 유해 곰팡이 검별용 학습모델 생성부를 더 포함하는 것을 특징으로 한다.

[0024] 또한 상기 실내 유해 곰팡이 검별 시스템은, 상기 생성한 곰팡이 유전체 염기서열 학습데이터를 활용하여 상기 사전 학습된 학습모델을 학습함에 있어서, 상기 사전 학습된 학습모델에서 특정 레이어의 사전 학습된 가중치 메트릭을 구성하는 가중치는 고정하며, 상기 특정 레이어의 사전 학습된 가중치 메트릭에서, 적어도 하나 이상의 소정의 가중치 메트릭을 선택하여 로우 랭크 디컴포지션(low rank decomposition)하고, 상기 학습시 누적 업데이트되는 그레디언트(gradient)에 따라 상기 로우 랭크 디컴포지션한 소정의 가중치 메트릭을 구성하는 가중치를 업데이트하여 상기 로우 랭크 디컴포지션한 소정의 가중치 메트릭을 곱한 다음, 상기 곱한 결과를 상기 가중치를 고정한 상기 선택한 소정의 가중치 메트릭에 합산함으로써 상기 사전 학습된 학습모델을 파인튜닝(fine tuning)하는 것을 더 포함하는 것을 특징으로 한다.

[0025] 또한 상기 유해 곰팡이 검별부는, 상기 유해 곰팡이 검별용 학습모델의 출력결과 중 확률이 제일 높은 레이블을 선택함으로써 유해 곰팡이를 검별하는 것을 포함하는 것을 특징으로 한다.

[0026] 또한 상기 실내 유해 곰팡이 검별 시스템은, 상기 수신한 곰팡이 유전체 염기서열이 소정의 길이 미만이면, 상기 길이에 맞게 0을 패딩하여 상기 수신한 곰팡이 유전체 염기서열을 전처리하는 전처리부를 더 포함하는 것을 특징으로 한다.

[0027] 아울러 본 발명의 일 실시예에 따른 실내 유해 곰팡이 검별 방법은, 곰팡이 유전체 염기서열 분석 장치로부터 실내 공기 중에서 포집된 곰팡이의 유전체 염기서열을 수신하는 곰팡이 유전체 염기서열 수신 단계 및 상기 수신한 곰팡이의 유전체 염기서열을 유해 곰팡이 검별용 학습모델에 입력하여 상기 곰팡이가 유해 곰팡이인지를 검별하는 유해 곰팡이 검별 단계를 포함하는 것을 특징으로 한다.

[0028] 또한 상기 실내 유해 곰팡이 검별 방법은, 사전에 수집한 곰팡이별 유전체 염기서열에 각 곰팡이가 독소를 생성하는지, 독성을 가지는지 혹은 인체에 무해한지에 대한 레이블을 각각 레이블링함으로써 곰팡이 유전체 염기서열 학습데이터를 생성하는 학습데이터 생성 단계 및 상기 생성한 학습데이터를 활용하여 상기 사전 학습된 학습모델을 학습시킴으로써 상기 유해 곰팡이 검별용 학습모델을 생성하는 유해 곰팡이 검별용 학습모델 생성 단계를 더 포함하는 것을 특징으로 한다.

[0029] 또한 상기 실내 유해 곰팡이 검별 방법은, 상기 생성한 곰팡이 유전체 염기서열 학습데이터를 활용하여 상기 사전 학습된 학습모델을 학습함에 있어서, 상기 사전 학습된 학습모델에서 특정 레이어의 사전 학습된 가중치 메트릭을 구성하는 가중치는 고정하며, 상기 특정 레이어의 사전 학습된 가중치 메트릭에서, 적어도 하나 이상의 소정의 가중치 메트릭을 선택하여 로우 랭크 디컴포지션(low rank decomposition)하고, 상기 학습시 누적 업데이트되는 그레디언트(gradient)에 따라 상기 로우 랭크 디컴포지션한 소정의 가중치 메트릭을 구성하는 가중치를 업데이트하여 상기 로우 랭크 디컴포지션한 소정의 가중치 메트릭을 곱한 다음, 상기 곱한 결과를 상기 가중치를 고정한 상기 선택한 소정의 가중치 메트릭에 합산함으로써 상기 사전 학습된 학습모델을 파인튜닝(fine tuning)하는 것을 더 포함하는 것을 특징으로 한다.

[0030] 또한 상기 유해 곰팡이 검별 단계는, 상기 유해 곰팡이 검별용 학습모델의 출력결과 중 확률이 제일 높은 레이블을 선택함으로써 유해 곰팡이를 검별하는 것을 포함하는 것을 특징으로 한다.

[0031] 또한 상기 실내 유해 곰팡이 검별 방법은, 상기 수신한 곰팡이 유전체 염기서열이 소정의 길이 미만이면, 상기 길이에 맞게 0을 패딩하여 상기 수신한 곰팡이 유전체 염기서열을 전처리하는 전처리 단계를 더 포함하는 것을 특징으로 한다.

발명의 효과

[0032] 이상에서와 같이 본 발명은 사전 학습된 학습모델을 파인튜닝하여 곰팡이 유전체 염기서열 학습데이터를 학습하여 유해 곰팡이 검별용 학습모델을 생성하고, 실제 실내 공기 중에서 포집한 곰팡이의 유전체 염기서열을 상기 유해 곰팡이 검별용 학습모델에 입력하여 유해 곰팡이를 신속하게 검별할 수 있는 효과가 있다.

[0033] 또한 본 발명은 사전 학습된 학습모델에서 특정 레이어의 가중치(가중치 메트릭)를 선택하여 파인튜닝하고, 상기 곰팡이 유전체 염기서열 학습데이터를 학습하는 과정에서 상기 선택한 특정 레이어의 가중치만 학습시킬 수 있도록 함으로써 학습시간 단축함과 동시에 유해 곰팡이 검별 정확도를 향상시킬 수 있는 효과가 있다.

[0034] 또한 본 발명은 유해 곰팡이 검별 결과를 제공하여, 유해 곰팡이 검별 결과에 따라 실내 공기를 신속하게 정화하여 유해 곰팡이로 인해 실내 거주자의 건강이 나빠지는 것을 방지할 수 있는 효과가 있다.

도면의 간단한 설명

[0035] 도 1은 본 발명의 일 실시예에 따른 실내 유해 곰팡이 검별 시스템 및 그 방법을 설명하기 위해 나타난 개념도이다.

도 2는 본 발명의 일 실시예에 따른 곰팡이 유전체 염기서열 학습데이터를 생성하는 방법을 설명하기 위해 나타난 도면이다.

도 3은 본 발명의 일 실시예에 따른 사전 학습된 학습모델을 설명하기 위해 나타난 도면이다.

도 4는 본 발명의 일 실시예에 따른 유해 곰팡이 검별용 학습모델을 설명하기 위해 나타난 도면이다.

도 5는 본 발명의 일 실시예에 따른 유해 곰팡이 검별 시스템의 구성을 나타낸 블록도이다.

도 6은 본 발명의 일 실시예에 따른 유해 곰팡이 검별용 학습모델을 생성하는 절차와, 유해 곰팡이를 검별하는 절차를 나타낸 흐름도이다.

도 7은 종래기술과 본 발명의 성능을 비교한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0036] 이하, 첨부한 도면을 참조하여 본 발명의 실내 유해 곰팡이 검별 시스템 및 그 방법에 대한 바람직한 실시예를 상세히 설명한다. 각 도면에 제시된 동일한 참조부호는 동일한 부재를 나타낸다. 또한 본 발명의 실시예들에 대해서 특정한 구조적 내지 기능적 설명들은 단지 본 발명에 따른 실시예를 설명하기 위한 목적으로 예시된 것으로, 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는 것이 바람직하다.

[0037] 도 1은 본 발명의 일 실시예에 따른 실내 유해 곰팡이 검별 시스템 및 그 방법을 설명하기 위해 나타난 개념도이다.

[0038] 도 1에 도시한 바와 같이, 본 발명의 일 실시예에 따른 실내 유해 곰팡이 검별 시스템(100)(이하, 유해 곰팡이 검별 시스템이라 칭함)은 실내 공기에서 포집한 곰팡이의 유전체 염기서열(단백질 서열)을 이용하여 유해 곰팡이를 검별하여 관리자 단말(200)로 검별 결과를 제공함으로써 실내 공기 정화를 신속히 수행할 수 있도록 한다.

[0039] 상기 곰팡이는 실내공간에 설치되는 곰팡이 포집 장치를 통해 포집된다. 상기 곰팡이 포집 장치는 사전에 설정한 주기에 따라 소정량의 실내공기를 수집함으로써 실내의 곰팡이를 포집하도록 구성될 수 있다. 상기 곰팡이 포집 장치는 KAS-110과 같은 미생물 공기샘플러로 구성될 수 있다.

[0040] 또한 포집된 곰팡이는 곰팡이 배양 장치를 통해 배양된다. 상기 곰팡이 배양 장치는 PDA 배지(potato dextrose agar plate), 스트렙토마이신(streptomycin)이 포함된 맥아 추출물 한천 배지, SD(sabouraud dextrose) 한천 배지 등과 같이 곰팡이 배양 배지를 이용하여 5일간 28도(섭씨)에서 포집된 곰팡이를 배양할 수 있다.

[0041] 또한 배양된 곰팡이는 곰팡이 유전체 염기서열 분석 장치를 통해 분석되어 곰팡이 유전체 염기서열이 추출된다.

[0042] 상기 곰팡이 유전체 염기서열 분석 장치는 ABI(applied biosystems) DNA 분석기와 같은 DNS 분석기로 구성될 수 있으며, 배양된 곰팡이(곰팡이균)가 함유된 증류수(100L)를 100도(섭씨)에서 10분간 가열한 후, 13,000rpm으로 5분간 원심분리 하여 불순물을 침전시킨 후 DNA 분석을 통해 곰팡이 유전체 염기서열을 추출한다.

[0043] 상기 곰팡이 유전체 염기서열 분석 장치는 추출한 곰팡이 유전체 염기서열을 유해 곰팡이 검별 시스템(100)으로 전송하도록 구성될 수 있다.

[0044] 이때, 곰팡이 유전체 염기서열은 ITS1(internal transcribed spacer 1) 및 ITS2(internal transcribed spacer 2)의 염기서열(nucleic acid sequence)로 구성된다.

[0045] 상기 유해 곰팡이 검별 시스템(100)은 실내 공기 중 유해 곰팡이를 검별하기 위해서 우선적으로 곰팡이별 유전

체 염기서열에 유해 곰팡이인지 혹은 인체에 무해한 곰팡이인지에 대한 레이블을 부가(레이블링)하여 곰팡이 유전체 염기서열 학습데이터를 생성하고, 상기 생성한 곰팡이 유전체 염기서열 학습데이터를 학습하여 유해 곰팡이 검별용 학습모델을 생성한다.

- [0046] 상기 학습은 사전 학습된 학습모델의 특정 레이어에서 선택한 가중치(더욱 자세하게는 가중치에 해당하는 학습 파라미터)만 학습할 수 있도록 파인튜닝(fine tuning)하여 수행된다.
- [0047] 상기 곰팡이 유전체 염기서열 학습데이터를 생성하는 것은 도 2를 참조하여 상세히 설명하도록 하며, 상기 학습은 도 3 및 도 4를 참조하여 상세히 설명하도록 한다.
- [0048] 상기 유해 곰팡이 검별 시스템(100)은 곰팡이 유전체 염기서열 분석 장치로부터 실제 실내공기에서 포집한 적어도 하나 이상의 곰팡이에 대한 곰팡이 유전체 염기서열이 수신되면, 유해 곰팡이 검별용 학습모델에 입력하여 해당 곰팡이가 유해 곰팡이인지 여부를 검별한다.
- [0049] 이후, 유해 곰팡이 검별 시스템(100)은 유해 곰팡이 검별 결과를 관리자 단말(200)로 전송하여 유해 곰팡이로 검별된 경우, 신속하게 실내 공기를 정화할 수 있도록 한다.
- [0050] 도 2는 본 발명의 일 실시예에 따른 곰팡이 유전체 염기서열 학습데이터를 생성하는 방법을 설명하기 위해 나타낸 도면이다.
- [0051] 도 2에 도시한 바와 같이, 본 발명의 일 실시예에 따른 유해 곰팡이 검별 시스템(100)은 학습을 위한 곰팡이별 유전체 염기서열을 수집하여 곰팡이 유전체 염기서열 학습데이터를 생성한다.
- [0052] 상기 학습을 위한 곰팡이별 유전체 염기서열은 UniProtKB(uniprot knowledge base)로부터 수집될 수 있으며, 상기 UniProtKB는 곰팡이별 유전체 염기서열과 각 곰팡이 유전체 염기서열에 곰팡이의 고유한 기능과 설명이 주석(description)으로 제공된다.
- [0053] 이에 따라 본 발명의 유해 곰팡이 검별 시스템(100)은 곰팡이별 유전체 염기서열의 주석을 참조하여 독성(virulent)이 포함되어 있으면 “독성”을 키워드로 하여 해당 곰팡이 유전체 염기서열에 레이블링하고, 독소(toxi)가 포함되어 있으면 “독소”를 키워드로 하여 해당 곰팡이 유전체 염기서열에 레이블링하며, 항독소(anti-toxi)를 포함하거나, 독성 및 독소를 포함하지 않으면 “정상(normal)”을 키워드로 하여 해당 곰팡이 유전체 염기서열에 레이블링함으로써 곰팡이 유전체 염기서열 학습데이터를 생성한다.
- [0054] 즉, 유해 곰팡이 검별 시스템(100)은 곰팡이가 독소를 생성하거나, 독성(즉, 독성으로 인해 사용자가 임상 증상이나 장애를 초래하는 경우)을 가지는 경우에는 “독소” 혹은 “독성”으로 레이블링하고, 이외 인체에 무해한 곰팡이의 경우에는 “정상”으로 레이블링함으로써 곰팡이 유전체 염기서열 학습데이터를 생성하는 것이다.
- [0055] 도 3은 본 발명의 일 실시예에 따른 사전 학습된 학습모델을 설명하기 위해 나타낸 도면이다.
- [0056] 도 3에 도시한 바와 같이, 본 발명의 일 실시예에 따른 사전 학습된 학습모델은 곰팡이 유전체 염기서열로부터 곰팡이의 종류를 분류(classification)하기 위해 사전 학습된 것이다.
- [0057] 상기 사전 학습된 학습모델은 ProtBERT로 구성될 수 있다.
- [0058] 상기 사전 학습된 학습모델은 ProtBERT로 구성될 수 있다.
- [0059] 상기 사전 학습된 학습모델은 피드포워드 네트워크(FFN, feed forward network)로 구성된 제1 레이어 및 멀티 헤드 어텐션(multi head attention) 네트워크로 구성된 제2 레이어를 포함하는 복수의 서브 레이어(예: 30개)를 포함하여 구성된다.
- [0060] 상기 멀티 헤드 어텐션 네트워크는 입력되는 곰팡이 유전체 염기서열에서 각 핵 염기에 Q(query), K(key), V(value)에 대한 가중치 매트릭(즉, W_q , W_k , W_v)(즉, Q, K, V에 대한 어텐션 헤드)을 곱하여 Q, K, V를 산출하고, 상기 Q, K, V를 scaled dot-product attention 연산을 통해 상기 곰팡이 유전체 염기서열에서 각 핵 염기 별로 다른 핵 염기에 대한 유사성 정보인 어텐션 결과를 출력한다.
- [0061] 상기 FFN은 뉴럴 네트워크(neural network)로 구성되며, 어텐션 결과를 입력으로 분류결과를 출력하게 된다.
- [0062] 도 4는 본 발명의 일 실시예에 따른 유해 곰팡이 검별용 학습모델을 설명하기 위해 나타낸 도면이다.
- [0063] 도 4에 도시한 바와 같이, 본 발명의 일 실시예에 따른 유해 곰팡이 검별용 학습모델은 곰팡이 유전체 염기서열 학습데이터를 이용하여 도 3을 참조하여 설명한 사전에 학습된 학습모델을 학습시킴으로써 생성된다.

- [0064] 상기 사전 학습된 학습모델은 모든 곱팡이의 종류에 대한 확률을 출력하도록 구성되므로, 모든 가중치를 사용할 필요가 없으며, 곱팡이 유전체 염기서열 학습데이터를 학습하여 곱팡이 유전체 염기서열에 해당하는 레이블에 대한 확률을 출력하도록 본 발명에 맞게 파인튜닝(fine tuning)해야 한다.
- [0065] 따라서 본 발명의 유해 곱팡이 검별 시스템(100)은 곱팡이 유전체 염기서열 학습데이터를 학습함에 있어서, 사전 학습된 학습모델에서 특정 레이어의 사전 학습된 가중치(W) 메트릭은 고정하고, 본 발명에 필요한 적어도 하나 이상의 소정의 특정 가중치 메트릭을 로우 랭크 디컴포지션하고, 상기 학습과정에서 누적 업데이트되는 그레디언트(gradient)에 따라 상기 로우 랭크 디컴포지션한 특정 가중치 메트릭을 구성하는 가중치를 업데이트하여 상기 로우 랭크 디컴포지션한 특정 가중치 메트릭을 곱해주고, 상기 곱한 결과를 상기 고정된 사전 학습된 가중치 메트릭에 합산할 수 있도록 하는 LoRA(low rank adaptation)를 적용하여 파인튜닝한다.
- [0066] 이때, 멀티 헤드 어텐션 네트워크는 합산 결과에 따른 어텐션 결과를 피드포워드 네트워크로 출력한다.
- [0067] 상기 특정 레이어는 멀티 헤드 어텐션 네트워크이며, 입력되는 곱팡이 유전체 염기서열을 포함하는 입력 임베딩(input embedding)에서 Q와 V를 생성하기 위해 사전 학습된 가중치 메트릭인 W_q 와 W_v 이다.
- [0068] 여기서 입력 임베딩은 곱팡이 유전체 염기서열에서 각 핵 염기의 위치(position)정보를 더 포함한다.
- [0069] 이렇게 파인튜닝하면 모든 학습파라미터를 사용하지 않고, 적은 수의 학습 파라미터로 학습을 수행할 수 있으며 학습모델을 경량화 함과 동시에 유해 곱팡이 검별을 위한 정확도를 향상시킬 수 있다.
- [0070] 상기 사전 학습된 학습모델을 파인튜닝한 본 발명의 유해 곱팡이 검별용 학습모델은 실제 곱팡이 유전체 염기서열이 입력되면 독소, 독성, 정상에 대한 레이블에 따른 확률을 출력하도록 구성된다.
- [0071] 입력된 곱팡이 유전체 염기서열을 구성하는 각 핵 염기간의 관련성 및 유사성을 판단하는 기준이 되는 것이 K이며, 상기 K는 사전학습모델에서 곱팡이 유전체 염기서열을 학습할 때 이미 학습되어 있다.
- [0072] 따라서 본 발명에서는 K에 대해서 이미 사전학습모델에서 곱팡이 유전체 염기서열을 학습하였으므로 새롭게 다시 학습할 필요가 없다. 즉, K에 대해서는 LoRA를 적용하지 않는다.
- [0073] 도 5는 본 발명의 일 실시예에 따른 유해 곱팡이 검별 시스템의 구성을 나타낸 블록도이다.
- [0074] 도 5에 도시한 바와 같이, 본 발명의 일 실시예에 따른 유해 곱팡이 검별 시스템(100)은 학습데이터 생성부(110), 파인튜닝부(120), 유해 곱팡이 검별용 학습모델 생성부(130), 곱팡이 유전체 염기서열 수신부(140), 전처리부(150) 및 유해 곱팡이 검별부(160)를 포함하여 구성된다.
- [0075] 상기 학습데이터 생성부(110)는 유해 곱팡이 검별용 학습 모델의 생성을 위한 곱팡이 유전체 염기서열 학습데이터를 생성한다.
- [0076] 상기 학습데이터 생성부(110)는 사전에 수집한 학습용 곱팡이별 유전체 염기서열에 곱팡이가 독소를 생성하는지, 독성을 가지는지 혹은 무해한지에 대한 레이블을 부가(레이블링)함으로써 곱팡이 유전체 염기서열 학습데이터를 생성한다.
- [0077] 이때, 곱팡이 유전체 염기서열의 최대 길이는 1024(즉, 유전체 염기서열을 구성하는 핵 염기의 개수)이며, 1024 미만이면 0으로 패딩된다.
- [0078] 상기 파인튜닝부(120)는 사전 학습된 학습모델을 LoRA를 적용함으로써 파인튜닝하여 곱팡이 유전체 염기서열 학습데이터를 학습할 수 있도록 한다.
- [0079] 상기 사전 학습된 학습모델은 곱팡이 유전체 염기서열을 학습하여 곱팡이의 종류를 분류하기 위한 것으로, 상기 파인튜닝은 도 4를 참조하여 설명하였으므로 여기서는 생략하도록 한다.
- [0080] 상기 유해 곱팡이 검별용 학습모델 생성부(130)는 학습데이터 생성부를 통해 생성한 곱팡이 유전체 염기서열 학습데이터를 이용하여 파인튜닝한 상기 사전 학습된 학습모델을 학습시켜 유해 곱팡이 검별용 학습모델을 생성한다.
- [0081] 상기 유해 곱팡이 검별용 학습모델의 입력은 실제 곱팡이의 곱팡이 유전체 염기서열이 되며, 출력은 각 레이블에 대한 확률이 된다.
- [0082] 상기 곱팡이 유전체 염기서열 수신부(140)는 곱팡이 유전체 염기서열 분석 장치로부터 실내 공기 중에서 포집한 곱팡이의 유전체 염기서열(즉, 곱팡이 유전체 염기서열)을 수신한다.

- [0083] 상기 전처리부는 수신한 곰팡이 유전체 염기서열이 사전에 설정한 길이(1024)보다 적으면 사전에 설정한 길이만큼 0으로 패딩하여 유해 곰팡이 검별용 학습모델의 입력에 적합하도록 전처리한다.
- [0084] 상기 유해 곰팡이 검별부(160)는 상기 전처리한 곰팡이 유전체 염기서열을 유해 곰팡이 검별용 학습모델에 입력하여 상기 곰팡이 유전체 염기서열에 해당하는 곰팡이가 인체에 유해한지를 검별한다.
- [0085] 상기 유해 곰팡이 검별부(160)는 유해 곰팡이 검별용 학습모델의 출력결과에 따라 상기 검별을 수행하는 것으로, 확률이 제일 높은 레이블을 선택함으로써 해당 곰팡이가 인체에 유해한지를 검별한다.
- [0086] 예를 들어, 유해 곰팡이 검별용 학습모델의 출력결과 독소 혹은 독성에 대한 레이블의 확률이 제일 높으면 유해 곰팡이 검별부(160)는 해당 곰팡이가 유해한 것으로 판단하고, 정상에 대한 레이블의 확률이 제일 높으면 상기 유해 곰팡이 검별부(160)는 해당 곰팡이가 무해한 것으로 판단함으로써 해당 곰팡이가 인체에 유해한지를 검별한다.
- [0087] 또한 상기 유해 곰팡이 검별 시스템(100)은, 상기 각 구성(110 내지 160)을 수행하는 프로세서를 포함하고, 상기 프로세서는 상기 유해 곰팡이 검별 방법을 수행하는 프로그램 코드를 실행하도록 구성되고, 상기 프로그램 코드는 메모리에 저장하도록 구성된다.
- [0088] 도 6은 본 발명의 일 실시예에 따른 유해 곰팡이 검별용 학습모델을 생성하는 절차와, 유해 곰팡이를 검별하는 절차를 나타낸 흐름도이다.
- [0089] 도 6에 도시한 바와 같이, 본 발명의 일 실시예에 따른 유해 곰팡이 검별용 학습모델을 생성하는 절차는 우선, 유해 곰팡이 검별 시스템(100)은 사전에 수집한 곰팡이별 유전체 염기서열을 기반으로 곰팡이 유전체 염기서열 학습데이터를 생성하는 학습데이터 생성 단계를 수행한다(S110).
- [0090] 상기 곰팡이 유전체 염기서열 학습데이터는 곰팡이별로 독소를 생성하는지, 독성을 가지는지 혹은 인체에 무해한지에 따른 레이블을 곰팡이별 유전체 염기서열에 레이블링함으로써 생성되는 상술한 바와 같다.
- [0091] 이때, 학습데이터 생성 단계는 사전에 설정한 길이 미만인 곰팡이 유전체 염기서열은 상기 길이만큼 0으로 패딩하여 곰팡이 유전체 염기서열 학습데이터를 생성하도록 한다.
- [0092] 다음으로 유해 곰팡이 검별 시스템(100)은 사전 학습된 학습모델을 파인튜닝하는 파인튜닝 단계를 수행한다(S120).
- [0093] 상기 파인튜닝은 사전 학습된 학습모델의 특정 레이어에 LoRA를 적용한 것으로, 도 4를 참조하여 설명하였으므로 여기서는 생략하도록 한다.
- [0094] 다음으로 유해 곰팡이 검별 시스템(100)은 파인튜닝한 사전 학습된 학습모델을 상기 생성한 곰팡이 유전체 염기서열 학습데이터를 활용하여 학습시킴으로써 유해 곰팡이 검별용 학습모델을 생성한다.
- [0095] 또한 본 발명의 일 실시예에 따른 실내 공기 중 유해 곰팡이를 검별하기 위한 절차는 우선, 유해 곰팡이 검별 시스템(100)은 곰팡이 유전체 염기서열을 수신하는 곰팡이 유전체 염기서열 수신 단계를 수행한다(S210).
- [0096] 상기 곰팡이 유전체 염기서열은 곰팡이 유전체 염기서열 분석 장치로부터 수신되거나, 혹은 관리자로부터 직접 입력될 수 있다.
- [0097] 다음으로 유해 곰팡이 검별 시스템(100)은 수신한 곰팡이 유전체 염기서열이 소정의 길이 미만이면 유해 곰팡이 검별용 학습모델에 적합한 데이터 포맷으로 변환하는 전처리 단계를 수행한다(S220).
- [0098] 상기 변환은 상기 수신한 곰팡이 유전체 염기서열이 소정의 길이 미만이면 해당 길이만큼 0으로 패딩함으로써 수행된다.
- [0099] 다음으로 유해 곰팡이 검별 시스템(100)은 전처리한 곰팡이 유전체 염기서열을 유해 곰팡이 검별용 학습모델에 입력하여 상기 전처리한 곰팡이 유전체 염기서열에 해당하는 곰팡이가 인체에 유해한지를 검별하는 유해 곰팡이 검별 단계를 수행한다(S230).
- [0100] 상기 유해 곰팡이 검별 시스템(100)은 유해 곰팡이 검별용 학습모델의 출력결과 중 확률이 제일 높은 레이블을 선택함으로써 유해 곰팡이를 검별한다.
- [0101] 다음으로 유해 곰팡이 검별 단계는 유해 곰팡이를 검별한 결과를 관리자 단말(200)로 전송하여, 유해 곰팡이 검별 결과에 따라 실내 공기를 신속히 정화할 수 있도록 한다.

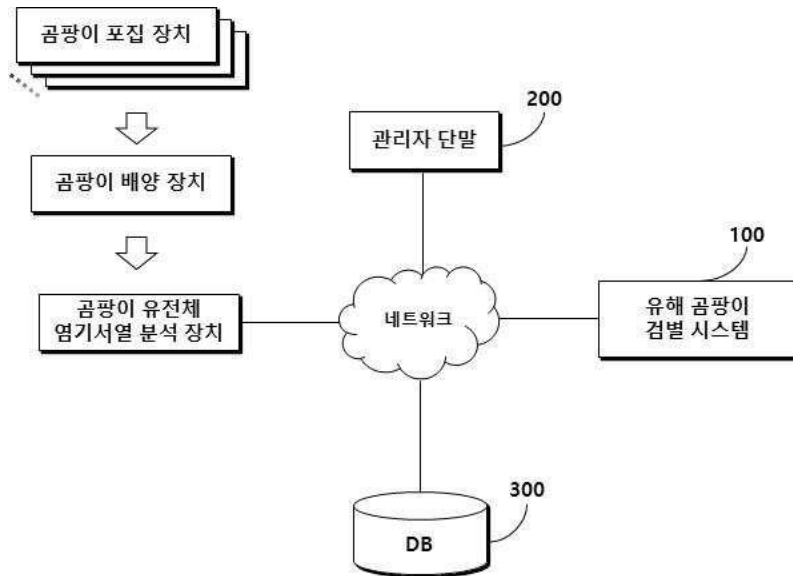
- [0102] 또한 상기 유해 곰팡이 검별 시스템(100)은, 상기 실내 유해 곰팡이 검별 방법을 수행하는 프로그램 코드를 저장하는 메모리(미도시) 및 상기 메모리에 저장된 상기 프로그램 코드를 실행하도록 구성된 프로세서(미도시)를 포함한다.
- [0103] 도 7은 종래기술과 본 발명의 성능을 비교한 도면이다.
- [0104] 도 7에 도시한 것과 같이, 사전 학습된 학습모델(ProtBERT)과 사전 학습된 학습모델을 LoRA를 적용하여 파인튜닝한 본 발명의 유해 곰팡이 검별용 학습모델의 성능을 비교한 것이다.
- [0105] 상기 성능을 비교하기 위해서 $1e-5$ 의 학습률로 20에포크(epochs) 동안 학습을 수행하였다.
- [0106] 본 발명의 유해 곰팡이 검별용 학습모델을 생성하기 위한 학습에서 훈련 가능한 매개변수는 983,040개이며 에포크당 평균 27분이 소요되었으며, 사전 학습된 학습모델을 통한 학습에서 훈련 가능한 매개변수는 420,983,811개이며 에포크당 평균 34분이 걸렸다. 이는, 본 발명이 적은 매개변수로 경량화하여 빠른 학습이 가능함을 알 수 있다.
- [0107] 또한 모델의 성능을 비교하기 위한 지표인 정확도(ACC, accuracy), 정밀도(precision)나 재현율(recall)을 나타내는 F1 점수, MCC(matthews correlation coefficient), auROC(area under receiver operating characteristic)를 이용하였다.
- [0108] 본 발명의 유해 곰팡이 검별용 학습모델은 사전 학습된 학습모델에 비해서, 테스트 세트에 대한 결과의 정확도(ACC)에서 0.0568, F1 점수에서 0.0555, MCC에서 0.0819, auROC에서 0.0166으로 모든 성능 측정 지표에서 증가되었음을 알 수 있다.
- [0109] 즉, 본 발명의 유해 곰팡이 검별용 학습모델은 사전 학습된 학습모델에 비해서 성능이 증가하였음을 알 수 있다.
- [0110] 이상에서 설명한 바와 같이 본 발명은 사전 학습된 학습모델을 LoRA를 적용하여 파인튜닝하여 곰팡이 유전체 염기서열 학습데이터를 학습함으로써 보다 정확하게 실내 공기 중에서 포집한 곰팡이 중 유해 곰팡이를 검별할 수 있는 효과가 있다.
- [0111] 아울러 이상에서는 본 발명의 바람직한 실시예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특정의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형 실시가 가능한 것은 물론이고, 이러한 변형 실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어서는 안 될 것이다.

부호의 설명

- [0112] 100: 유해 곰팡이 검별 시스템
 110: 학습데이터 생성부
 120: 파인튜닝부
 130: 유해 곰팡이 검별용 학습모델 생성부
 140: 유해 곰팡이 염기서열 수신부
 150: 전처리부
 160: 유해 곰팡이 검별부

도면

도면1



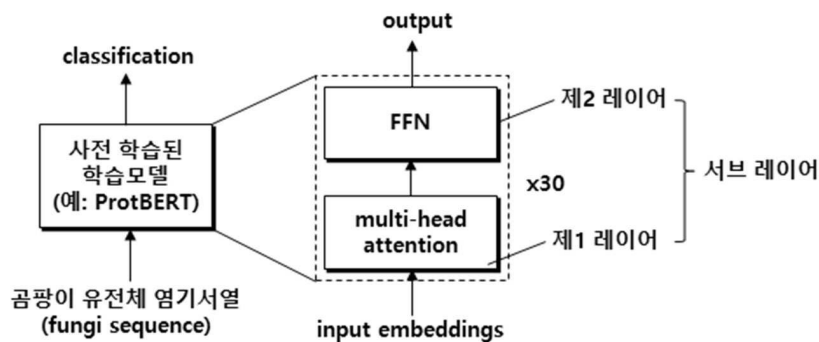
도면2

Algorithm Pseudo Code of Data Collection(곰팡이 유전체 염기서열 학습데이터 생성)

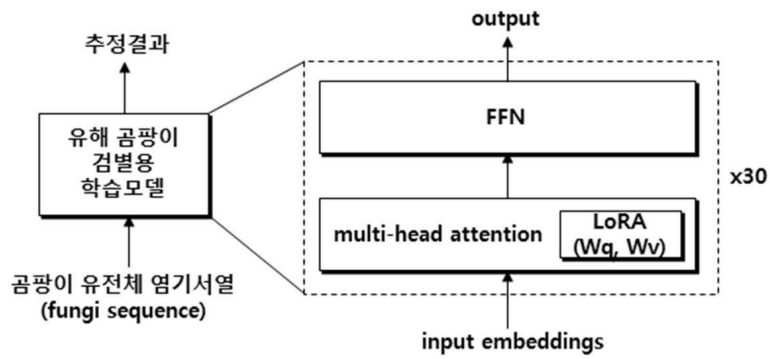
```

for protein sequences(곰팡이 유전체 염기서열) do
  if length of protein sequences <= 1024 then
    if description contains virulent then
      label(key) sequences as "virulent"
    else if description contains toxi then
      if description contains anti-toxin then
        label sequences as "normal"
      else
        label sequences as "toxin"
      end if
    else
      label sequence as "normal"
    end if
  end if
end if
  
```

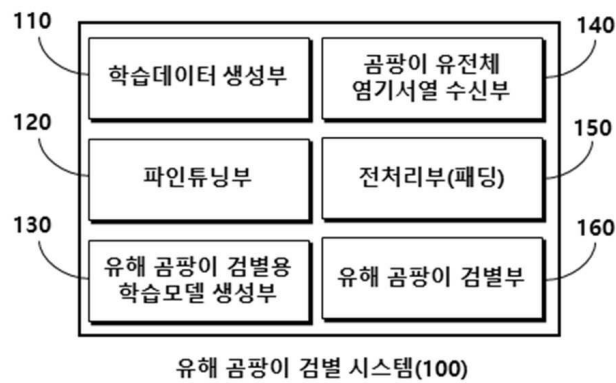
도면3



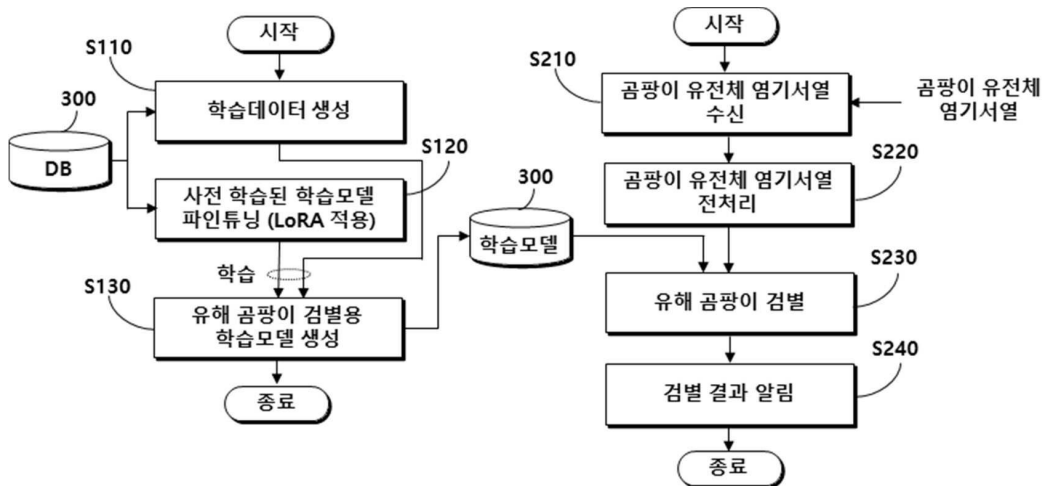
도면4



도면5



도면6



도면7

Method		ACC	F1	MCC	auROC
ProtBERT	Valid	0.8330	0.8345	0.7547	0.9325
	Test	0.8377	0.8389	0.7599	0.9297
ProtBERT with LoRA (본 발명)	Valid	0.9096	0.9098	0.8645	0.9554
	Test	0.8942	0.8944	0.8418	0.9463