

아이디어 공모전

시뮬레이션을 활용한
멀티 오믹스 데이터 분석 및
인공지능 신약개발 활용

목차

Index

⋮

01

제안이유

02

제안내용

03

활용 방안
및 기대효과

04

참고문헌

⋮

01 제안이유

기존 모델의 한계

- 1 오믹스 데이터의 복잡성과 방대한 양으로 인해 의미 있는 정보를 추출하고 필수적인 의생명학적 중요성을 파악하는 데 어려움이 존재
- 2 데이터의 차원이 증가함에 따라 기존의 표 형식 데이터에서 중요한 연관성 표현이 어려움
- 3 데이터를 활용해서 질병 (발생) 상태를 분류하는 다양한 인공지능 기반 분석 모형이 개발되어 왔지만, 복잡한 자연 과정을 설명하는 분석 모형은 해석 가능성이 부족한 '블랙박스' 모델인 경우가 빈번함

01 제안이유

멀티-오믹스 분석 프로그램

배경

오믹스 분야에서 처리량이 많은 시퀀싱 기술의 발전으로 데이터 폭증
풍부한 데이터는 정밀 의학을 위한 예측 모델링의 기회를 제공했지만,
데이터 분석과 결과 해석의 과제 존재
기존 기계학습 기술은 오믹스 데이터 기반 예측 모델을 생성하지만
데이터 내의 잠재적 관계 처리에 한계가 있음.

제안

보다 정확한 예측을 위해서는 기존 기계 학습 기술의 한계를 극복할
수 있는 새로운 접근법이 요구됨
오믹스 데이터 간의 복잡한 상호작용을 모델링하고,
이를 통해 의생명학 의미를 해석할 수 있는 통합적인 분석 체계 제시
오믹스 데이터 뿐만 아니라, ADMET (Absorption, Distribution,
Metabolism, Excretion, and Toxicity) 고차원 데이터 분석에 활용 가능

01 제안이유

시뮬레이션을 활용한
멀티 오믹스 데이터 분석
프로그램

1

테이블 형식 오믹스 데이터를 이미지와 같은 표현으로 변환
=> CNN(:Convolutional Neural Network) 등 딥러닝 모델을
활용하여 잠재적 관계 후보 판별

2

전이 학습 활용으로 예측력 향상 뿐만 아니라 계산 시간
감소와 분 정석확도 향상 기대

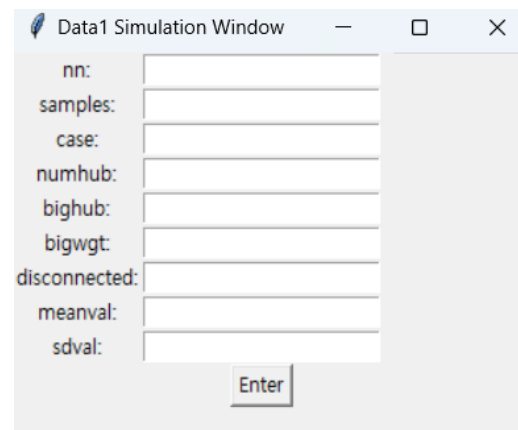
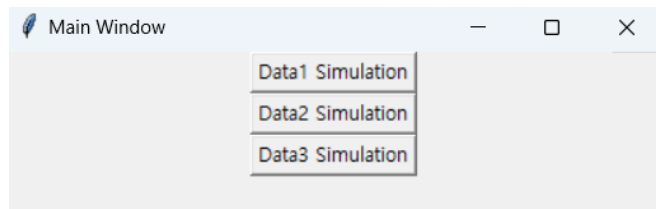
3

CNN을 활용한 오믹스 데이터의 예측 모델링은
기존 기계 학습 기술의 한계 극복과 혁신적 성과 달성 기대

02 제안내용

데이터 시뮬레이션

- 다양한 분포와 상황을 가정하여 데이터를 생성함으로써 모델의 일반화 성능 향상
- 실제 데이터 수집의 어려움이나 한계를 극복하여 충분한 양의 데이터 확보
- 특정 시나리오에 대한 모델 성능 사전 테스트로 모델 개발 과정 효율화



[다양한 상황을 가정할 수 있는 데이터 시뮬레이션 프로그램과 옵션 윈도우]

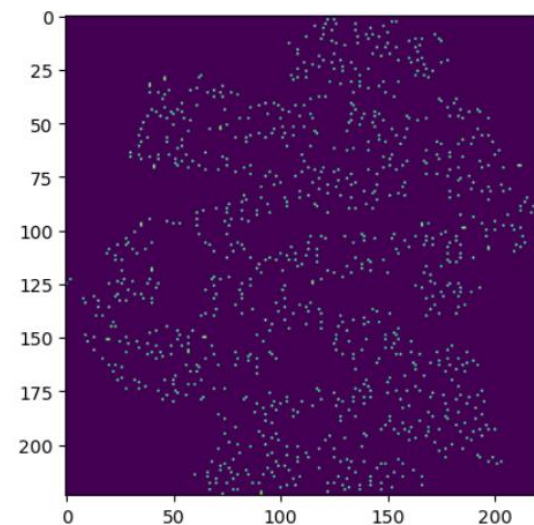
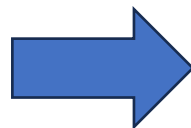
02 제안내용

이미지 데이터로의 표현

```
a["simudata"][1]
```

```
array([ -0.         ,  0.         ,  0.         ,  0.         ,  
        0.45036076, -0.         , -0.         , -0.         ,  
        0.         , -0.         , -0.         ,  0.         ,  
        0.         ,  0.         ,  0.         ,  0.1749668 ,  
       -0.         ,  0.         , -0.         ,  0.         ,  
        0.         , -0.59856086,  0.         ,  0.20804128,  
        0.         , -0.         ,  0.         , -0.87392217,  
       0.35861055, -0.39910944, -0.         ,  0.         ,  
        0.         , -0.51800457,  0.         , -0.         ,  
       -0.         ,  0.16057246, -0.         , -0.79830721,  
      -0.07304266, -0.         ,  0.         , -0.         ,  
       0.67910423, -0.         ,  0.         , -0.         ,  
       -0.         ,  0.         ,  0.         , -0.         ,  
      69.6313113 ,  0.         ,  0.         ,  0.         ,  
        0.         ,  0.         , -0.         ,  0.         ,  
      20.3081136 , -0.         , -0.         ,  0.         ,
```

[시뮬레이션으로 생성된 표 형식 데이터]



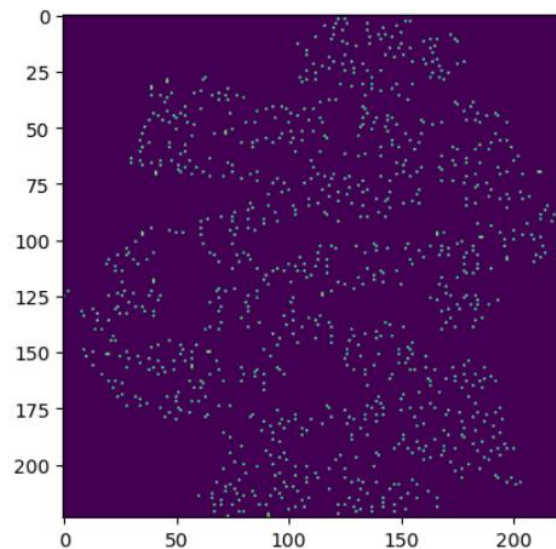
[이미지로 표현된 시뮬레이션 데이터]

오믹스 데이터는 대부분 변수가 서로 독립적으로 표시되는 " 표 형식 " 데이터에서 나오기 때문에 오믹스 데이터 간의 잠재적인 관계파악에 효과적이지 못함

이미지 분석 기법을 오믹스 데이터에 적합하게 적용하여 오믹스 데이터 간의 잠재적인 관계를 보다 잘 파악할 수 있음

02 제안내용

이미지 데이터로의 표현



[이미지로 표현된 시뮬레이션 데이터]

이미지 데이터의 재구성을 통해, 유사한 특성을 공유하는 요소를 근접한 이웃 특성으로 배열하는 반면, 근접하지 않은 요소는 현 상태를 유지하면서, CNN이 다각적이고 효율적으로 적용될 수 있는 풍부한 환경을 생성

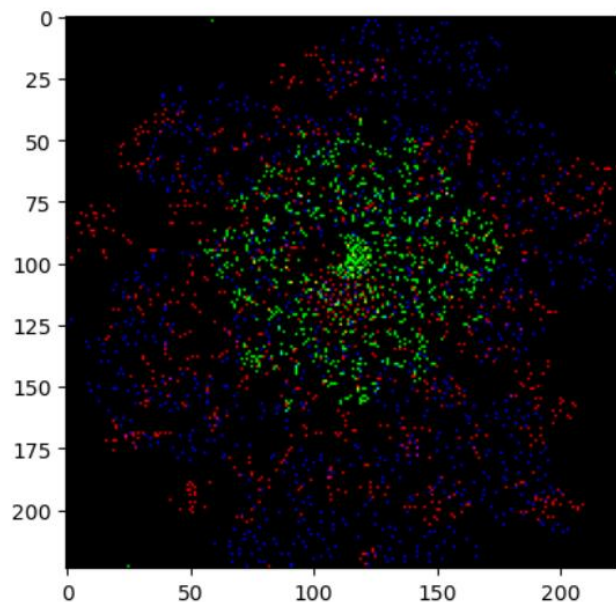
=> 유사한 변수를 가까이 모아 하나의 그룹으로 처리하여 오믹스 데이터의 복잡한 구조와 패턴을 반영

02 제안내용

이미지 데이터로의 표현

```
# np.stack을 사용하여 RGB 이미지 생성
rgb_image = np.stack((image1, image2, image3), axis=-1)

# 이미지 시각화
plt.imshow(rgb_image)
plt.show()
```



[이미지로 표현된 멀티오믹스 데이터]

오믹스 데이터는 유전체, 전사체, 단백질체, 대사체 등 개별 생물학적 층위를 다루지만, 생물학적 시스템은 이러한 층위들이 복잡하게 상호작용하며 작동

시뮬레이션으로 얻은 오믹스 데이터들의 정보를 통합하여 표현한 하나의 이미지를 딥러닝을 통해 다양한 오믹스 데이터들 간의 시너지 상호 작용을 포착하여 전체적인 이해를 촉진하고 분석을 위한 보다 풍부한 내용을 제공

멀티오믹스 데이터 분석을 통해 질병의 다양한 층위에서의 변화를 통합적으로 파악할 수 있기 때문에, 질병 기전을 보다 심도 있게 규명 가능함

02 제안내용

이미지데이터 분석을 위한 딥러닝



딥러닝을 이용한
멀티오믹스데이터의
관계추론

- 1 수많은 상호 작용을 식별하고 비선형 효과를 모델링
=> 과적합 위험을 완화할 수 있는 효과적인 정규화 기능 제공
- 2 오믹스 데이터의 다양한 유형의 구조화된 정보 수용
- 3 이질적인 데이터를 공동으로 분석하고 통합하여 보다 포괄적인 이해 가능

전이학습

- 딥러닝 모델 구축에는 많은 자원이 필요
- 전이학습은 대규모 데이터 세트의 지식을 재사용하여 소규모 데이터에서도 성능 향상 가능
- 사전 훈련된 모델을 새로운 작업에 맞게 특성화하면서 데이터 수집 필요성을 줄이면서도 성능 향상

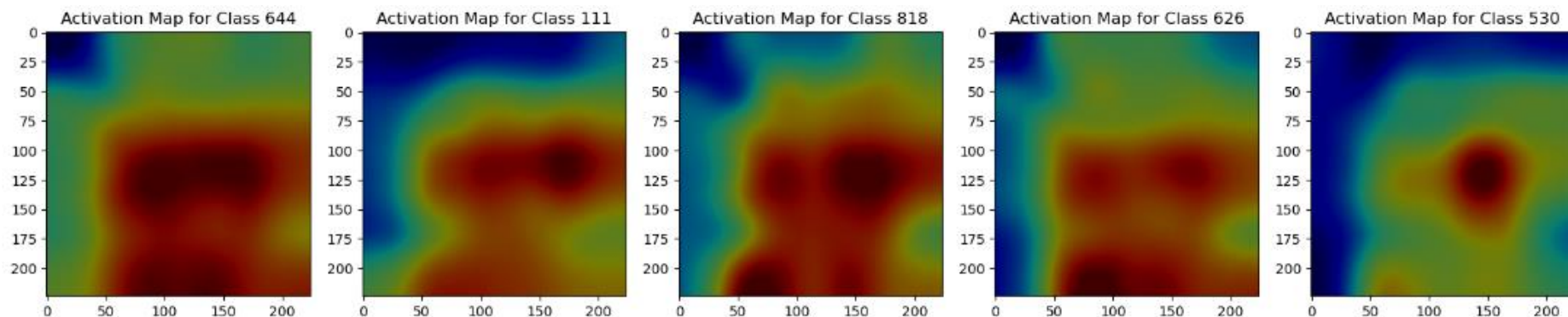
데이터 시뮬레이션과 전이학습

- 전이학습은 대규모 데이터의 지식을 활용하여 소규모 데이터에서도 성능 향상
- 데이터 시뮬레이션은 부족한 실제 데이터를 인공적으로 생성하여 모델 학습 지원
- 두 기술을 함께 활용하면 제한적인 데이터에서도 강력한 딥러닝 모델 구축 가능

02 제안내용

이미지데이터 분석을 위한 딥러닝

관심영역을 보여주는 CAM



[각 클래스 별로 예측에 영향을 미치는 이미지 내 영역 강조 표시]

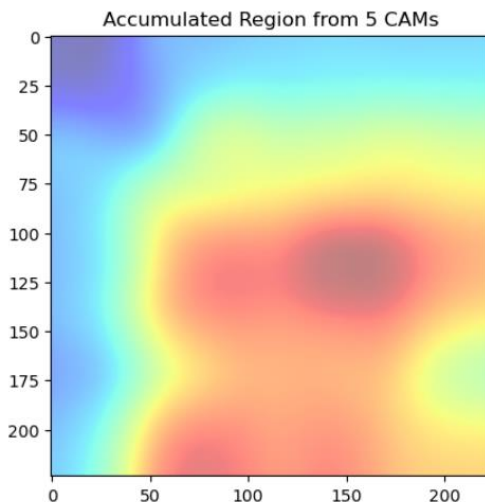
CAM(:Class Activation Map)을 통해 각 클래스(예: 질병 유무, 치료 반응 등)에 대해 모델이 예측할 때 큰 영향을 미치는 이미지의 특정 영역을 강조 표시

=> 중요한 생물학적 층위나 변수들을 식별할 수 있어 모델의 예측 과정을 해석할 수 있게 도와주기 때문에 모델의 성능을 개선하거나, 모델의 신뢰성을 높일 수 있음

02 제안내용

이미지데이터 분석을 위한 딥러닝

관심영역을 보여주는 CAM



[모든 클래스에 공통적으로 영향을 미치는 영역 강조하여 표시]

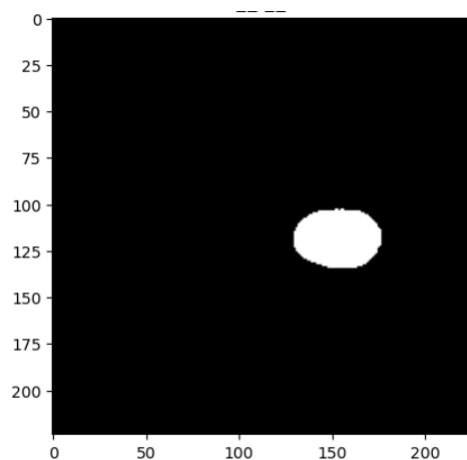
모든 클래스에서 공통적으로 활성화되는 멀티오믹스 데이터의 영역을 살펴보는 누적 CAM을 통해 복잡한 생물학적 네트워크와 상호작용을 통합적으로 이해할 수 있음

클래스별 CAM과 누적 CAM 결과를 원본 멀티오믹스 데이터와 연계하여 분석하면, 질병 발생, 치료 반응 등의 생물학적 메커니즘을 보다 심도 있게 이해할 수 있음

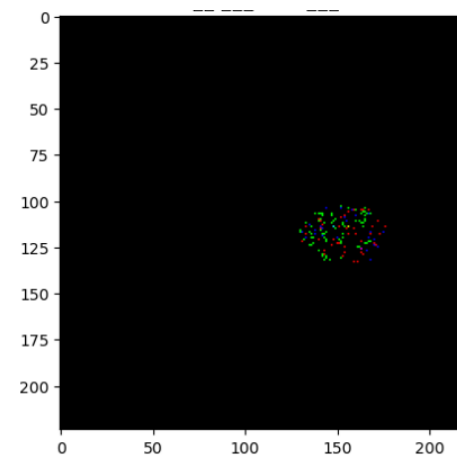
02 제안내용

이미지데이터 분석을 위한 딥러닝

관심영역을 보여주는 CAM



[CAM을 통해 얻은 이미지의 관심 영역]

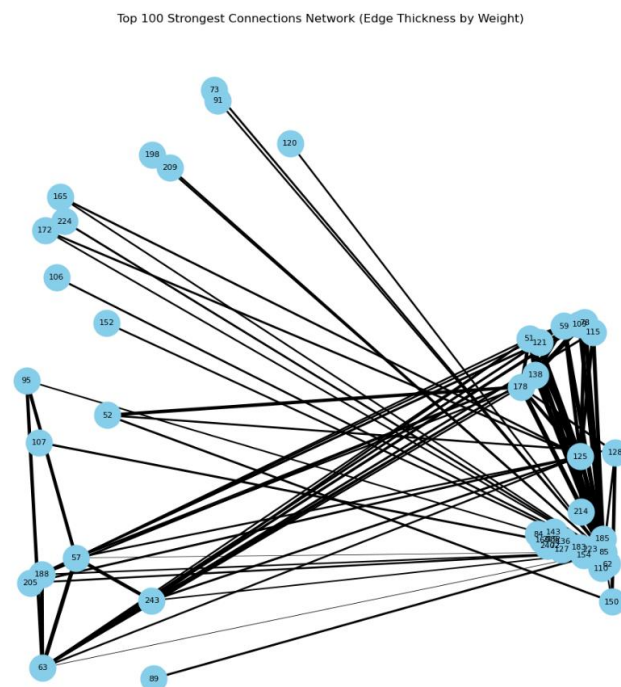


[관심영역 내에 위치한 원본 멀티오믹스 데이터]

CAM을 통해 모델이 주목하는 입력 데이터의 관심 영역에 해당하는 원본 멀티오믹스 데이터를 분석하여 모델이 중요하게 여기는 데이터 간 상호작용 및 관계 파악 가능

02 제안내용

네트워크 분석을 통한 인사이트 도출



[관심영역 변수들 간 주요 네트워크]

관심 영역의 변수들 간 상관관계, 네트워크 구조 등을 분석하여 생물학적 시스템 내 변수 간 상호작용 이해

=> 질병 발생, 치료 반응 등의 복잡한 생물학적 과정을 보다 구체적인 설명가능

딥러닝을 활용한 모델의 생물학적 해석능력 향상

03 활용 방안 및 기대 효과

데이터 시뮬레이션의 기대 효과

1

실제 데이터 수집에 소요되는 비용과 시간 절감

2

다양한 시나리오에 대한 모델 성능 평가를 통해 최적의 모델 선별을 통한 개발 비용 절감

3

시뮬레이션을 통해 도출된 모델의 실제 환경 적용 시 초기 투자 비용 절감

03 활용 방안 및 기대 효과

생물학적 시스템 이해 증진

- 유전체, 전사체, 단백질체, 대사체 등 다양한 오믹스 데이터를 통합적으로 분석을 통한 생물학적 시스템의 복잡한 상호작용 이해

- 시뮬레이션을 통해 실험적으로 검증하기 어려운 가설에 대한 테스트 및 검증 가능

생물학적 시스템 최적화

- 시뮬레이션을 통해 생물학적 시스템의 특정 매개변수를 조절하여 사용자에게 맞춤형 경험 제공

- 다양한 시나리오에 대한 가정과 실험은 유전자 조작, 대사 경로 설계, 약물 개발 등에 활용 가능

03 활용 방안 및 기대 효과

개인 맞춤형 의료 및 건강 관리

- 개인의 오믹스 데이터를 분석하고 시뮬레이션하여 질병 예방, 치료, 관리 등에 활용할 수 있음
- 개인의 유전적 특성과 생활 습관을 고려한 맞춤형 건강관리 솔루션 제공

교육 프로그램의 역할

- 생물학 교육에서 복잡한 생물학적 시스템을 시각화하고 이해를 돕는 도구로 활용 가능
- 생물학 연구 분야에서 새로운 가설 검증, 실험 설계, 데이터 분석 등에 활용 가능

04 참고문헌

Alok Sharma, Artem Lysenko, Shangru Jia, Keith A. Boroevich & Tatsuhiko Tsunoda, Advances in AI and machine learning for predictive medicine, Journal of Human Genetics, 2024

Alok Sharma, Edwin Vans, Daichi Shigemizu, Keith A. Boroevich & Tatsuhiko Tsunoda, DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture, Scientific Reports , 2019

Anastasia Zompola, Aigli Korfiati, Konstantinos Theofilatos, and Seferina Mavroudid, Omics-CNN: A comprehensive pipeline for predictive analytics in quantitative omics using one-dimensional convolutional neural networks, Heliyon, 2023