

## 【서지사항】

【서류명】	특허출원서
【참조번호】	P20240519KR
【출원구분】	특허출원
【출원인】	
【명칭】	연세대학교 산학협력단
【특허고객번호】	2-2005-009509-9
【대리인】	
【명칭】	특허법인 지담
【대리인번호】	9-2018-100261-1
【지정된변리사】	김재홍
【포괄위임등록번호】	2023-020263-8
【발명의 국문명칭】	DDPM 및 그래프 주의 네트워크 기반의 인공지능 약물 반응 성 예측 시스템 및 그 예측 방법
【발명의 영문명칭】	DDPM AND GRAPH ATTENTION NETWORK-BASED AI DRUG RESPONSE PREDICTION SYSTEM AND METHOD THEREOF
【발명자】	
【성명】	박상현
【성명의 영문표기】	PARK, Sang Hyun
【주민등록번호】	670101-1XXXXXX
【우편번호】	03626
【주소】	서울특별시 서대문구 세무서8길, 101동 1502호
【발명자】	

【성명】	최승연
【성명의 영문표기】	CHOI, Seung Yeon
【주민등록번호】	970828-1XXXXXX
【우편번호】	07213
【주소】	서울특별시 영등포구 당산로45길 7-3, 101동 808호
【발명자】	
【성명】	서상민
【성명의 영문표기】	SEO, Sang Min
【주민등록번호】	930507-1XXXXXX
【우편번호】	63272
【주소】	제주특별자치도 제주시 고마로 44, 801호
【출원언어】	국어
【심사청구】	청구
【공지예외적용대상증명서류의 내용】	
【공개형태】	논문발표(학회지공개)
【공개일자】	2024. 10. 19
【이 발명을 지원한 국가연구개발사업】	
【과제고유번호】	1711198526
【과제번호】	00229822
【부처명】	과학기술정보통신부
【과제관리(전문)기관명】	한국연구재단
【연구사업명】	원천기술개발사업

**【연구과제명】** [통합이지바로/주관] 난치성 질환 극복을 위한 인공지능 기반의 다중 약물 적응증 최적화 플랫폼 개발 및 혁신신약 발굴 (1/2단계)(1/2)

**【과제수행기관명】** 연세대학교

**【연구기간】** 2024.01.01 ~ 2024.12.31

**【취지】** 위와 같이 특허청장에게 제출합니다.

대리인 특허법인 지담 (서명 또는 인)

### 【수수료】

**【출원료】** 0 면 46,000 원

**【가산출원료】** 34 면 0 원

**【우선권주장료】** 0 건 0 원

**【심사청구료】** 12 항 778,000 원

**【합계】** 824,000원

**【감면사유】** 전담조직(50%감면)[1]

**【감면후 수수료】** 412,000 원

**【첨부서류】** 1. 공지예외적용대상(신규성상실의예외, 출원시의특례)규정을 적용받기 위한 증명서류\_1통

1 : 공지예외적용대상(신규성상실의예외, 출원시의특례)규정을 적용받기 위한 증명서류

[PDF 파일 첨부](#)

## 【발명의 설명】

### 【발명의 명칭】

DDPM 및 그래프 주의 네트워크 기반의 인공지능 약물 반응성 예측 시스템 및 그 예측 방법{DDPM AND GRAPH ATTENTION NETWORK-BASED AI DRUG RESPONSE PREDICTION SYSTEM AND METHOD THEREOF}

### 【기술분야】

【0001】 본 발명은 DDPM 및 그래프 주의 네트워크 기반의 인공지능 약물 반응성 예측 시스템 및 그 예측 방법에 관한 것으로, 더욱 상세하게는 생물학적 경로 분석을 통한 유전자 특징을 활용한 데이터 증강 및 약물 반응 예측 모델을 수립하여 고차원 데이터를 효과적으로 처리하고 예측 성능을 향상시키는 약물 반응 예측 시스템에 관한 것이다.

### 【발명의 배경이 되는 기술】

【0003】 약물 반응성 예측은 신약 개발 및 환자의 유전적 특성에 맞는 효과적인 치료법을 찾는 것을 목표로 하는 현대 정밀 의학에서 매우 중요한 요소이다.

【0004】 동일한 종류의 암을 진단받고 그 진행도 비슷한 것으로 파악되는 환자들은 동일한 치료 방법을 적용하더라도 환자들마다 다른 효과가 관찰되고 있기 때문이다.

【0005】 유사한 치료 방법과 처방을 적용하더라도 환자마다 다른 치료 효과를 보이거나 부작용이 발생하는 등 치료에 대한 반응성이 다른 이유는, 개인마다 다른 독특한 유전적 특성이 가장 큰 요인으로 알려져 있다.

【0006】 이에 환자의 유전적 특성을 고려한 치료 방법이 연구될 필요가 있으며, 최근 유전체 연구가 활발해지면서 환자의 유전적 특성을 분석하는 연구가 계속되고 있다.

【0007】 도 1은 종래의 기술에 따라 환자 고유의 약물 반응성을 예측하는 약물 반응 예측 시스템을 도시하는 도면이다.

【0008】 종래의 약물 반응 예측 시스템(101)은, 환자의 게놈 데이터 및 다양한 유전자 정보가 포함된 오믹스 데이터(11)를 기계 학습 모델(21)의 학습 데이터로 이용하였다.

【0009】 또한, 다양한 기계 학습 방식을 복수 통합하거나, 응용함으로서 환자 고유의 약물 반응성을 보다 정밀히 예측하도록 하는 모델이 연구되고 있다.

【0010】 그러나, 종래의 예측 모델은 데이터의 본질적인 특성으로 인해 예측 효능을 저해하는 문제가 발생한다. 특히 생물학적 데이터는 높은 차원성의 정보를 지니고 있으며, 샘플 수가 적으면 정확한 예측 모델을 훈련하기 어렵다.

【0011】 유전자 데이터의 경우, 샘플 수가 제한되어 있으며, 유전자 간 복잡한 관계로 인해 데이터 분석이 더욱 복잡해지고 모델 훈련 중 과적합이라는 위험 요소가 발생한다.

【0012】 이러한 문제를 극복하기 위해, 유전자 데이터를 통한 인공지능 훈련 방법에 대해 실제 유전자 발현 프로파일과 유사한 새로운 데이터 포인트를 시뮬레이션할 필요가 있다. 또한, 약물 반응성 예측에 생물학적 정보가 반영됨으로써, 복잡한 유전자 상호 작용을 포착할 수 있는 예측 방법이 필요한 설정이다.

### 【선행기술문헌】

#### 【특허문헌】

【0013】 (특허문헌 0001) 한국공개특허 10-2024-0079608호(2024.06.05)

#### 【발명의 내용】

#### 【해결하고자 하는 과제】

【0014】 본 발명이 해결하고자 하는 기술적 과제는 인공지능을 이용하여 약물 반응성을 예측하는 수단을 제공하는 것이며, 제한된 데이터를 활용함으로써 예측 정확도가 높은 인공지능 약물 반응성 예측 시스템 및 그 방법을 제공하는 것이다.

【0015】 본 발명이 해결하고자 하는 기술적 과제는 데이터의 고차원적인 특성이 감소되도록 처리하여 컴퓨팅 연산에 있어 과도한 부하를 방지하고 데이터 과적합을 감소시켜 예측 효율을 높이는 인공지능 약물 반응성 예측 시스템 및 그 방법을 제공하는 것이다.

【0016】 본 발명이 이루고자 하는 기술적 과제는 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

### 【과제의 해결 수단】

【0017】 상기 기술적 과제를 달성하기 위하여, 본 발명의 일실시예에 따른 DDPM 및 그래프 주의 네트워크 기반의 인공지능 약물 반응성 예측 시스템은, 환자의 유전자 데이터를 수집하는 유전자 데이터 입력부와, 타겟 유전자와 생물학적 경로 유전자 간의 근접성을 계산하여 중요한 유전자를 선별하는 유전자 데이터 전처리부와, 전처리된 유전자 데이터를 입력받아 DDPM을 통해 원본 데이터와 유사한 증강 데이터를 생성하고, 그래프 주의 네트워크를 통해 약물 반응성을 예측하는 약물 반응성 예측 인공지능 모델 및 예측된 약물 반응성을 출력하는 약물 반응성 출력부를 포함한다.

【0018】 상기 기술적 과제를 달성하기 위하여, 본 발명의 일실시예에 따른 DDPM 및 그래프 주의 네트워크 기반의 인공지능 약물 반응성 예측 방법은, 환자의 유전자 데이터를 수집하는 단계와, 타겟 유전자와 경로 유전자 간의 근접성을 계산하여 중요한 유전자를 선별하는 단계와, 선별된 유전자 데이터를 DDPM(Denoising Diffusion Probabilistic Models)을 이용해 증강 데이터를 생성하는 단계와, 증강된 데이터를 그래프 주의 네트워크를 통해 약물 반응성을 예측하는 단계 및 예측된

약물 반응성을 출력하는 단계를 포함한다.

### 【발명의 효과】

【0019】 본 발명의 실시예에 따르면, 환자 유전자 데이터를 이용하여 학습한 인공지능 약물 반응성을 예측하는 시스템을 제공함으로써, 환자 개인별 치료를 위한 신약 개발에 대해 정확성과 신뢰성을 최적화할 수 있다.

【0020】 본 발명의 실시예에 따르면, 인공지능 약물 반응성 예측 시스템은 데이터 증강을 수행하고 증강된 데이터를 학습에 이용함으로써, 학습 데이터의 부족 없이 약물 반응성 예측 정확도를 높일 수 있다.

【0021】 또한, 본 발명의 실시예에 따르면, 인공지능 약물 반응성 예측 시스템은 데이터의 고차원적인 특성이 감소되도록 처리하여 컴퓨팅 연산에 있어 과도한 부하를 방지하고 데이터 과적합 발생률을 감소시킬 수 있다.

【0022】 본 발명의 효과는 상기한 효과로 한정되는 것은 아니며, 본 발명의 상세한 설명 또는 특허청구범위에 기재된 발명의 구성으로부터 추론 가능한 모든 효과를 포함하는 것으로 이해되어야 한다.

### 【도면의 간단한 설명】

【0024】 도 1은 종래 기술의 약물 반응성 예측 시스템(101)을 도시한 도면이다.

도 2는 본 발명의 실시예에 따른 인공지능 약물 반응성 예측 시스템(100)을

도시한 도면이다.

도 3은 본 발명의 실시예에 따른 약물 반응성 예측 인공지능 모델 수립 과정을 도시한 흐름도이다.

도 4는 본 발명의 실시예에 따른 유전자 데이터 전처리 과정을 도시한 흐름도이다.

도 5는 본 발명의 실시예에 따른 약물 반응성 예측 인공지능 모델(130)의 DDPM 수행 과정을 도시한 흐름도이다.

도 6은 본 발명의 실시예에 따른 약물 반응성 예측 인공지능 모델(130)의 약물 반응성 예측 과정을 도시한 흐름도이다.

도 7은 본 발명의 실시예에 따른 약물 반응성 예측 인공지능 모델(130)의 예측 성능을 도시한 도면이다.

도 8은 본 발명의 실시예에 따른 약물 반응성 예측 방법을 도시한 흐름도이다.

도 9는 본 발명의 실시예에 따른 약물 반응성 예측 시스템(100)을 구현하는 컴퓨팅 장치를 도시한다.

### 【발명을 실시하기 위한 구체적인 내용】

【0025】 이하에서는 첨부한 도면을 참조하여 본 발명을 설명하기로 한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 따라서 여기에서 설명하는 실시예로 한정되는 것은 아니다. 그리고 도면에서 본 발명을 명확하게 설명

하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

【0026】 명세서 전체에서, 어떤 부분이 다른 부분과 "연결(접속, 접촉, 결합)"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 부재를 사이에 두고 "간접적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 구비할 수 있다는 것을 의미한다.

【0027】 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다.

【0028】 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

【0029】 이하 첨부된 도면을 참고하여 본 발명의 실시예를 상세히 설명하기로 한다.

【0031】 도 2는 본 발명의 실시예에 따른 DDPM 및 그래프 주의 네트워크 기반의 인공지능 약물 반응성 예측 시스템(100)을 도시한 도면이다.

【0032】 도 2를 참조하면 DDPM 및 그래프 주의 네트워크 기반의 인공지능 약물 반응성 예측 시스템(100)은, 유전자 데이터 입력부(110), 유전자 데이터 전처리부(120), 약물 반응성 예측 인공지능 모델(130), 약물 반응성 출력부(140)를 포함한다.

【0033】 먼저, 유전자 데이터 입력부(110)는, 암세포주 정보 데이터, GE, 암세포주와 약물 화합물 간의 반응성 정보 데이터 등 환자 유전자 데이터를 입수할 수 있다.

【0034】 이때, 유전자 데이터 입력부(110)는 상기 데이터를 GDSC(Genomics of Drug Sensitivity in Cancer), TCGA(The Cancer Genome Atlas) 및 PDX(Patient-Derived Xenograft) 와 같은 데이터베이스 중 적어도 하나에 접속하여 입수할 수 있다.

【0035】 한편, DDPM 및 그래프 주의 네트워크 기반의 인공지능 약물 반응성 예측 시스템에서 이용하는 GE(gene expression data)는, 유전자 데이터로 기재하기로 한다.

【0036】 유전자 데이터 전처리부(120)는, 유전자 데이터 입력부(110)에서 입수된 데이터를 약물 반응성 예측 모델의 입력 데이터로 가공 처리할 수 있다.

【0037】 유전자 데이터 전처리부(120)는 약물 관련 유전자와 경로 유전자 간의 생물학적 경로를 식별하여 중요 유전자를 선별할 수 있다.

【0038】 구체적으로, 유전자 데이터 전처리부(120)는 대상 약물과 가장 밀접하게 관련된 생물학적 경로를 식별하여 대상 약물과 관련성이 높은 생물학적 경로를 구성하는 유전자를 선별할 수 있다.

【0039】 예를 들어, 입력된 유전자 데이터를 분석하여, 약물 반응성 예측에 중요한 유전자 집합을 도출한다. 이 과정에서 유전자 데이터 전처리부(120)는 유전자 간 생물학적 경로를 식별하고 타겟 유전자와 관련된 경로 유전자 간의 통계적 상관성을 계산하여 주요 경로를 구성하는 유전자들을 선별한다. 예를 들어, Z-점수를 사용하여 생물학적 경로의 유전자들의 통계적 유의성을 평가한 후, 예측에 중요한 유전자들로 최종 유전자 집합을 형성하게 된다. 이와 같이, 유전자 데이터 전처리부(120)는 중요 유전자만을 선별해냄으로서, 데이터가 지닌 많은 정보로부터 발생하는 고차원성 문제를 해결할 수 있다.

【0040】 약물 반응성 예측 인공지능 모델(130)은, 유전자 데이터 전처리부(120)에서 생성된 약물 반응 관련 유전자 데이터를 입력으로 기계 학습하여 인공신경망 기반의 약물 반응성 예측 모델을 수립할 수 있다.

【0041】 약물 반응성 예측 인공지능 모델(130)은, DDPM과 그래프 주의 네트워크 기반의 알고리즘을 통해 예측 모델을 학습한다. 먼저, 전처리된 유전자 데이터를 DDPM 기반의 데이터 증강 모듈에 입력하여 원본 데이터와 유사한 증강 데이터

를 생성한다. 이를 위해 그래프 오토인코더(GAE)를 사용해 유전자 데이터의 생물학적 관계를 반영한 저차원 잠재 공간을 생성하고, 이를 DDPM 학습에 활용하여 고차원 데이터의 희소성 문제를 해결한다. 증강된 데이터는 그래프 주의 네트워크를 통해 노드(유전자) 간의 관계 패턴을 학습하여 약물 반응성을 예측한다

【0042】 여기서 수립되는 약물 반응성 예측 인공지능 모델(130)은 도 5 내지 6에서 후술하기로 한다.

【0043】 약물 반응성 출력부(140)는, 약물 반응성 예측 인공지능 모델(130)을 통해 예측된 결과를 출력한다. 여기서 출력되는 약물 반응성은 약물 반응성 출력부(140)에 코드 및 프로그램으로서 구비될 수도 있지만, 별도의 출력 모듈을 통해 출력될 수 있다.

【0044】 출력된 약물 반응성 결과는 제약 연구원이나 의료 기관에 제공될 수 있으며, 사용자 요청에 따라 그래프나 차트 형태로 시작화되어 전달될 수도 있다.

【0045】 도 3은 본 발명의 실시예에 따른 약물 반응성 예측 인공지능 모델 수립 과정을 도시한 흐름도이다.

【0046】 여기서는 선별된 중요 경로 데이터를 입력으로 그래프 AE(그래프 오토인코더 (GAE)) 및 DDPM(Denoising Diffusion Probabilistic Models (DDPM))을 이용하여 학습을 수행할 수 있다.

【0047】 그래프 AE는 기계학습에 있어 차원 압축을 수행할 수 있는 신경망이다. 그래프AE는 인코더를 통해 차원 압축을 수행함으로서, DDPM과 같은 모델이 보

다 작은 차원 공간에서 실행되어 모델 학습에 필요한 계산 리소스를 줄일 수 있다.

【0048】 그래프 AE와 DDPM을 모두 활용함으로써, 유전자 발현 데이터를 효과적으로 생성시키고 훈련 데이터 세트를 증강할 수 있다. 이와 함께 결과로서 약물 반응 예측 모델의 일반화 가능성을 향상시켜 예측 효율을 높일 수 있다.

【0049】 도3을 참조하면, 단계(S110)에서 유전자 데이터 입력부(S110)는 약물 반응성 예측 모델의 학습을 위한 유전자 데이터를 입수한다. 상기 유전자 데이터는 GDSC(Genomics of Drug Sensitivity in Cancer), TCGA(The Cancer Genome Atlas) 및 PDX(Patient-Derived Xenograft) 와 같은 데이터베이스 중 적어도 하나에 접속하여 입수할 수 있다.

【0050】 단계(S120)에서 유전자 데이터 전처리부(120)는 유전자의 생물학적 경로를 식별하여 중요 경로 유전자 데이터를 선별할 수 있다.

【0051】 단계(S120)에서 유전자 데이터 전처리부(120)는 약물 관련 유전자와 경로 유전자 간의 근접성을 측정하여 관련 생물학적 경로를 식별하여 예측 결과에 가장 영향력 있는 유전자를 선별할 수 있다. 구체적으로, 유전자 데이터 전처리부(120)는 대상 약물과 가장 밀접하게 관련된 생물학적 경로를 식별하여 대상 약물과 관련성이 높은 생물학적 경로를 구성하는 유전자를 선별할 수 있다.

【0052】 예를 들어, 입력된 유전자 데이터를 분석하여, 약물 반응성 예측에 중요한 유전자 집합을 도출한다. 이 과정에서 유전자 데이터 전처리부(120)는 유전자 간 생물학적 경로를 식별하고 타겟 유전자와 관련된 경로 유전자 간의 통계적

상관성을 계산하여 주요 경로를 구성하는 유전자들을 선별한다. 예를 들어, Z-점수를 사용하여 생물학적 경로의 유전자들의 통계적 유의성을 평가한 후, 예측에 중요한 유전자들로 최종 유전자 집합을 형성하게 된다.

【0053】 단계(S130)에서는, 단계(S120)에서 전처리된 유전자 데이터를 기반으로 DDPM(Denoising Diffusion Probabilistic Model) 기반의 생성 모델을 통해 증강 데이터를 생성한다. 먼저, 그래프AE(GAE, graph auto-encoder)를 이용하여 유전자 데이터의 고차원 정보를 저차원 잡재 공간으로 압축한 후, DDPM을 통해 원본 데이터와 유사한 추가 데이터 포인트를 생성한다. 증강된 데이터는 학습 데이터의 희소성을 극복하여 모델의 예측 성능을 향상시키며, DDPM을 활용한 증강 데이터는 유전자 간의 생물학적 관계를 유지함으로써 데이터의 다양성을 확보한다

【0054】 또한, 단계(S130)에서 DDPM 생성 모델을 통해 원본과 유사한 증강 데이터를 생성하여 학습 데이터로 이용할 수 있다.

【0055】 단계(S140)에서는 약물 반응성 예측 인공지능 모델에 의해 약물 반응성을 예측하고 예측된 결과를 제공할 수 있다. 여기서, 그래프 주의 네트워크 기반의 예측 모델이 구축된다. 그래프 주의 네트워크는 각 유전자 간의 상호작용을 고려하여 그래프 상의 노드 간 중요도를 학습하며, 노드(유전자) 정보의 집계와 결합을 통해 약물 반응성 예측 결과를 생성한다. 최종 예측 결과는 약물 반응성 출력부에서 코드 형태로 제공되며, 사용자가 필요로 할 경우 그래프나 차트 형태로 시각화하여 제공될 수 있다

【0056】 도 4는 본 발명의 실시예에 따른 유전자 데이터 전처리 과정을 도시

한 흐름도이다.

【0057】 먼저, 단계(S121)에서, 유전자 데이터 전처리부(120)는 입수된 유전자 데이터로부터 유전자 집합을 구성한다. 도 4를 참조하면, 대상 약물(m)이 존재하며 대상 약물(m)과 가장 관련성이 있는 타겟 단백질(t), 경로 유전자(s1, s2, ...) 설정을 확인할 수 있다.

【0058】 단계(S122)에서, 유전자 데이터 전처리부(120)는 약물과 관련된 생물학적 경로를 식별할 수 있다. 구체적으로, 유전자 데이터 전처리부(120)는 대상 약물(m)과 관련된 타겟 단백질(t)과 각 경로 유전자(s1, s2, ...) 간의 경로를 식별할 수 있다. 이를 위해 유전자 데이터 전처리부(120)는 타겟 유전자(t)로부터 각 경로 유전자(s1, s2, ...)까지의 거리를 측정할 수 있으며, 다음의 계산식(수식1)을 이용할 수 있다.

【0059】 수식 1.

$$d_{closest} = \frac{1}{|T|} \sum_{t \in T} \min_{s \in S} d(s, t)$$

【0060】

【0061】 T: 타겟, s:경로 유전자, d(s, t):약물 관련 유전자와 경로 유전자 간의 최단 경로.

【0062】 단계(S123)에서, 유전자 데이터 전처리부(120)는 경로 유전자 집합들로부터 기준 분포를 생성할 수 있다.

【0063】 단계(S123)에서 유전자 데이터 전처리부(120)는 각 경로의 계산된 거리가 경로를 구성하는 노드(유전자)의 수와 상관없이 통계적으로 유의미한지 여부를 확인하기 위해 랜덤으로 분포를 생성하고, 각 경로의 거리에 대한 기준 분포의 평균과 표준 편차를 사용하여 z-점수를 계산한다.

【0064】 단계(S124)에서, 유전자 데이터 전처리부(120)는 유전자 선별을 위한 통계적 관련성 지표를 도출할 수 있다.

【0065】 예를 들어, 약물과 가장 밀접한 관련성을 가진 경로 유전자는 가장 짧은 거리 경로의 유전자인 것으로 관련성 지표가 도출되었다.

【0066】 단계(S125)에서, 유전자 데이터 전처리부(120)는 통계적 관련성 지표를 적용하여 중요한 유전자를 선택할 수 있다.

【0067】 도 4를 참조하면, (s1), (s2)가 타겟 유전자(t)와 가까운 경로 유전자이며, 이 경로에 포함된 유전자들이 선택된다.

【0068】 이와 같이 중요한 유전자를 선택하는 데이터 차원 축소 과정을 통해, 약물 반응성 예측 시스템(100)은 연산 컴퓨팅의 과부화 및 데이터의 과적합 문제를 해소할 수 있다.

【0069】 도 5는 본 발명의 실시예에 따른 약물 반응성 예측 인공지능 모델(130)의 DDPM 수행 과정을 도시한 흐름도이다.

【0070】 DDPM(Denoising Diffusion Probabilistic Model)은 생성형 인공지능 모델 중 하나로서, 약물 반응성 예측 인공지능 모델(130)의 부족한 학습 데이터로

인한 일반화 부족 문제를 해결하도록 데이터 증강을 수행할 수 있다.

【0071】 여기서 DDPM 은 학습된 그래프 AE(graph auto-encoder)를 이용하여 유전자 데이터 학습 차원을 축소시킴으로서 학습에 필요한 계산을 감축할 수 있다.

【0072】 도 5를 참조하면, 단계(S131)에서, 약물 반응성 예측 인공지능 모델(130)은 선별된 유전자 데이터를 그래프 AE에 입력한다.

【0073】 여기서 유전자 데이터 전처리부(120)가 선별한 유전자 데이터가 입력된다.

【0074】 단계(S132)에서, 약물 반응성 예측 인공지능 모델(130)은 상기 입력 데이터를 그래프AE를 통해 인코딩하여 학습 차원을 축소할 수 있다.

【0075】 단계(S133)에서, 약물 반응성 예측 인공지능 모델(130)은 차원 축소된 데이터 입력층에 DDPM 학습을 실행한다.

【0076】 단계(S134)에서, 약물 반응성 예측 인공지능 모델(130)은 DDPM 생성 학습에 의해 잠재변수를 생성할 수 있다.

【0077】 단계(S135)에서, 약물 반응성 예측 인공지능 모델(130)은 생성된 잠재변수를 그래프AE 디코더를 이용하여 원본 유전자 데이터 수준으로 변환할 수 있다.

【0078】 단계(S136)에서, 약물 반응성 예측 인공지능 모델(130)은 단계(S135)에서 변환된 데이터로 증강 데이터를 생성할 수 있다.

【0079】 또한, 약물 반응성 예측 인공지능 모델(130)은 단계S134 ~ 단계S135를 반복 수행하여 증강 데이터를 생성시킨다.

【0080】 이상의 도5를 참조하여 설명된 프로세스는 다음과 같은 학습 알고리즘을 적용하여 수행될 수 있다.

---

**Algorithm 1:** Overall data augmentation process

---

```

1  Input
2   $x \leftarrow$  Gene expression data
3   $c \leftarrow$  Response labels for the records in  $x$ 
4   $\mathbb{G} \leftarrow$  Graph information
5   $\psi, \phi \leftarrow$  Weights for the AE
6   $\theta \leftarrow$  Weights for the generative (diffusion-based) model
7  repeat
8    Take gradient descent step at
9     $\nabla_\psi \|x - De_\phi(En_\psi(x, \mathbb{G}))\|^2$ 
10    $\nabla_\phi \|x - De_\phi(En_\psi(x, \mathbb{G}))\|^2$ 
11  until converged
12 repeat
13   $(x_o, c) \sim p(x, c)$ 
14   $r_o \leftarrow En_\psi(x_o, A)$ 
15   $t \sim \text{Uniform}(\{1, \dots, T\})$ 
16   $\epsilon \sim \mathcal{N}(0, \mathbb{I})$ 
17  Take gradient descent step at
18   $\nabla_\theta \|\epsilon - \epsilon_\theta(\sqrt{\bar{a}_t}r_o + \sqrt{1 - \bar{a}_t}\epsilon, t, c)\|^2$ 
19  until converged
20   $r_T \sim \mathcal{N}(0, \mathbb{I})$ 
21  for  $t = T, \dots, 1$  do
22     $z \sim \mathcal{N}(0, \mathbb{I})$  if  $t > 1$ , else  $z = 0$ 
23     $r_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( r_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_t}} \epsilon_\theta(r_t, t, c) \right) + \sigma_t z$ 
24  end for
25  return  $r_o$ 
26  Augmented sample  $\leftarrow De_\phi(r_o)$ 

```

---

【0081】

【0082】 이와 같이 약물 반응성 예측 모델(130)은 그래프 AE와 DDPM을 모두 활용하여 유전자 데이터를 효과적으로 증강함으로써 모델의 일반화를 달성하고 예측 성능을 향상시킬 수 있다.

【0084】 도 6은 본 발명의 실시예에 따른 약물 반응성 예측 인공지능 모델

(130)의 약물 반응성 예측 과정을 도시한 흐름도이다.

【0085】 도 6을 참조하면, 약물 반응성 예측 인공지능 모델(130)은 그래프

주의 네트워크를 이용하여 약물 반응성을 예측할 수 있다.

【0086】 그래프 주의 네트워크는 데이터의 그래프 구조에서 그래프의 각 노드에 대해 각각의 중요도 정보를 반영할 수 있는 신경망으로서 유전자 간의 관계 패턴을 학습하여 중요도를 도출할 수 있다.

【0087】 먼저, 단계(S141)에서 약물 반응성 예측 인공지능 모델(130)은 환자의 유전자 데이터를 취득하여 그래프 구조화할 수 있다.

【0088】 여기서, 환자의 유전자 데이터는 DDPM 증강에 의해 생성된 데이터 및 공개된 데이터베이스로부터 취득하는 GE(Gene expression) 데이터를 포함한다.

【0089】 상기 유전자 데이터의 생물학적 경로는 하위 그래프에 표현되며, 상기 하위 그래프는 노드 특징과 노드 간 상호 작용을 설명하는 엣지로 구성될 수 있다.

【0090】 여기서 노드 특징은 유전자 발현 및 유전자 지표로 이루어지고, 유전자 지표는 학습 가능한 임베딩 행렬을 이용하여 할당된다.

【0091】 즉, 유전자 지표를 학습 가능한 유전자 임베딩 행렬을 이용하여 추출하였으므로, 생물학적 경로에서 공유된 정보를 활용하면서도 각 경로의 고유한 특성을 학습할 수 있게 된다.

【0092】 단계(S142)에서 약물 반응성 예측 인공지능 모델(130)은 유전자 연결 정보를 부가하여 그래프 관계 학습을 수행할 수 있다.

【0093】 여기서, 약물 반응성 예측 모델에 부가되는 유전자 연결 정보는 사전 학습을 위한 것으로, KEGG(Kyoto Encyclopedia of Genes and Genomes) 데이터베이스로부터 취득될 수 있다.

【0094】 또한, 상기 그래프 관계 학습은 기존의 학습된 그래프 주의 신경망 모델을 통해 수행될 수 있다.

【0095】 단계(S143)에서 약물 반응성 예측 인공지능 모델(130)은 하위 그래프의 정보를 집계하여 노드 정보를 결합한다.

【0096】 여기서, 다음 수식과 같이 리드아웃 함수를 이용하여 하위 그래프 정보를 집계하고 해당 하위 그래프를 나타내는 생물학적 경로 및 타겟 단백질 간의 거리 정보를 이 하위 그래프의 노드 정보와 결합하는 연산을 수행한다.

【0097】 수식 2.

$$Z_G = concat(d_k^{-1} * MLP_{readout}(H^{L,k}) \mid k = 0, \dots, K)$$

【0099】 여기서  $d_k$ 는 타겟 단백질과 해당 하위 그래프(생물학적 경로) 간의 거리를 나타내며,  $H^{L,k}$ 는 k 번째 하위 그래프의 최종 노드 상태를 나타낸다.

【0100】 수식 2와 같이 타겟 단백질과의 거리 정보를 리드아웃 함수에 포함 시킴으로써, 약물 반응성 예측 모델은 유전자 발현과 약물 반응 간의 관계를 더 잘 포착하고 보다 정확한 예측을 수행하게 된다. 이는 타겟 단백질과 가까운 유전자가

더 멀리 있는 유전자보다 약물 반응에 더 큰 영향을 미칠 수 있기 때문이다.

【0101】 단계(S144)에서 약물 반응성 예측 인공지능 모델(130)은 각 노드 정보 결합 후 하나의 표현 벡터로 집결한다. 즉, 약물 반응성 예측 인공지능 모델(130)은 신경망에서의 정보가 모두 결합된 정보를 하나의 표현 벡터로서 생성할 수 있다(수식3).

【0102】 수식 3.

$$[0103] \quad y_{pred} = MLP_{pred}(Z_G)$$

【0104】 그리고 단계(S145)에서 약물 반응성 예측 인공지능 모델(130)은 단계(S144)의 표현 벡터에 기초하여 약물 반응 예측 결과를 도출한다.

【0105】 도 7은 본 발명의 실시예에 따른 약물 반응성 예측 인공지능 모델(130)의 예측 성능을 도시한 도면이다.

【0106】 도 7을 참조하면, 예측 성능을 설명하기 위한 실험 결과표를 확인할 수 있다. 여기서 실험 결과표는 다양한 기준 모델과 제안된 모델의 약물 반응 예측에 대한 일반화 성능을 개관한 것이다.

【0107】 실험 결과표에 의하면, 약물 반응성 예측 인공지능 모델(130)의 실험 모델(DiffGA)이 PDX 벤치마크 데이터셋에서 모든 약물에 대해 가장 높은 AUC 값을 얻었으며, erlotinib를 제외한 모든 약물에서 최고의 성능을 기록한 결과(r)를 확인할 수 있다.

【0108】 또한, TCGA 벤치마크 데이터셋이 모든 약물에 대해 최고의 성능을 나타내어 비교 대상 기준 모델을 능가했다는 것을 나타냈으며, 6개 데이터셋에 대한 평균 AUC 값이 0.79로 나타났다. 이 결과(r)를 통해 제안된 방법이 PDX 및 TCGA 벤치마크 데이터셋에서 약물 반응 예측의 정확성을 향상시키는 것임을 확인되었다.

【0109】 도 8은 본 발명의 실시예에 따른 약물 반응성 예측 방법을 도시한 흐름도이다.

【0110】 먼저, 단계(S151)에서 약물 반응성 예측 시스템(100)은, 환자의 유전자 데이터를 취득한다. 여기서 취득되는 환자의 유전자 데이터는 GE(Gene expression) 데이터를 포함하는 것이며, 공개된 데이터베이스에 접속하여 취득할 수 있다.

【0111】 단계(S152)에서 약물 반응성 예측 시스템(100)은 생물학적 경로 유전자 데이터를 선택한다.

【0112】 단계(S153)에서 약물 반응성 예측 시스템(100)은 생물학적 경로에 기초한 유전자 데이터를 입력으로 기계 학습을 수행한다.

【0113】 단계(S154)에서 약물 반응성 예측 시스템(100)은 약물 반응성 예측 인공지능 모델을 이용하여 약물 반응성 예측을 수행한다.

【0114】 단계(S155)에서 약물 반응성 예측 시스템(100)은 약물 반응성에 대한 예측을 출력한다. 출력된 약물 반응성 예측 결과는 예를 들어 치료 기관 및 신약 개발 기관의 사용자에게 제공될 수 있으며, 출력된 예측 결과에 기반하여 제약

및 환자 치료에 이용될 수 있다.

【0115】 도 9는 본 발명의 실시예에 따른 약물 반응성 예측 시스템(100)을 구현하는 컴퓨팅 장치를 도시한다.

【0116】 도 2 내지 도 8에 의해 설명된 본 발명의 실시에는 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치(900)로 구현될 수 있다.

【0117】 컴퓨팅 장치(900)는 프로세서(910), 메모리(920), 스토리지(930), 통신 인터페이스(940), 시스템 인터커넥트(950) 및 디스플레이(960)을 포함할 수 있다.

【0118】 프로세서(910)는, CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit), GPU(Graphic Processing Unit) 및 APU(Application Processing Unit)을 포함한다.

【0119】 메모리(920)는 프로세서(910)와 상호작용하여 프로그램이 효율적으로 실행될 수 있도록 데이터를 저장하고 필요한 정보에 빠르게 접근할 수 있도록 하는 기능을 수행한다. 메모리(920)는 레지스터, 캐시 메모리, 주 메모리, 읽기 전용 메모리, 가상 메모리, 비휘발성 메모리 중 적어도 하나를 포함한다.

【0120】 스토리지(930)는 데이터를 영구적으로 저장하고 관리하는 역할을 한다. 스토리지는 컴퓨팅 시스템이 꺼지거나 재부팅된 후에도 데이터를 보존하며, 운영 체제, 애플리케이션, 사용자 파일 등을 저장하는 데 사용된다. 스토리지(930)은, 하드 디스크 드라이브(HDD), 솔리드 스테이트 드라이브(SSD), 광학 디스크, 네

트워크 스토리지 및 클라우드 스토리지 중 적어도 하나를 포함한다.

【0121】 통신 인터페이스(940)는 컴퓨팅 시스템 내부 및 외부의 다양한 장치들 간에 데이터를 주고받기 위한 경로를 제공한다. 통신 인터페이스(940)는 USB(Universal Serial Bus), PCIe(Peripheral Component Interconnect Express), SATA(Serial ATA), Ethernet, Wi-Fi, Thunderbolt 및 HDMI(High-Definition Multimedia Interface) 중 적어도 하나의 통신 방식을 지원할 수 있다.

【0122】 시스템 인터커넥트(950)는 컴퓨팅 시스템 내부에서 다양한 구성 요소들 간의 데이터와 신호를 주고받는 역할을 한다. 시스템 인터커넥트(950)는, 버스(Bus), 포인트-투-포인트(Point-to-Point), 크로스바 스위치(Crossbar Switch), 네트워크-온-칩(Network-on-Chip, NoC) 중 적어도 하나의 방식을 지원할 수 있다.

【0123】 디스플레이(960)는 컴퓨팅 시스템의 출력 장치로서, 사용자에게 시각적인 정보를 제공하는 기능을 수행한다.

【0124】 전술한 구성에 의하여, 본 발명의 실시예에 따른 프로그램은, 프로세서(910)에 의해 실행되는 명령어들에 기초하여 실행되며, 메모리(920) 또는 스토리지(930)에 저장될 수 있다.

【0126】 전술한 본 발명의 실시예에 따른 방법은 다양한 컴퓨터 구성요소를 통하여 실행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능한 기록매체에 기록될 수 있다. 컴퓨터 판독 가능한 기록매체는 프로그램 명령어, 데이터 파일

일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능한 기록매체에 기록되는 프로그램 명령은 본 발명의 실시예를 위하여 특별히 설계되고 구성된 것이거나, 컴퓨터 소프트웨어 분야의 통상의 기술자에게 공지되어 사용가능한 것일 수 있다. 컴퓨터 판독 가능한 기록매체는, 하드디스크, 플로피디스크, 자기테이프 등의 자기기록 매체, CD-ROM, DVD 등의 광기록 매체, 플롭티컬디스크 등의 자기-광 매체, ROM, RAM, 플래시 메모리 등과 같이, 프로그램 명령을 저장하고 수행하도록 구성된 하드웨어를 포함한다. 프로그램 명령은, 컴파일러에 의해 만들어지는 기계어 코드, 인터프리터를 사용하여 컴퓨터에서 실행될 수 있는 고급 언어 코드를 포함한다. 하드웨어는 본 발명에 따른 방법을 처리하기 위하여 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있고, 그 역도 마찬가지이다.

**【0127】** 본 발명의 실시예에 따른 방법은 프로그램 명령 형태로 전자장치에서 실행될 수 있다. 전자장치는 스마트폰이나 스마트패드 등의 휴대용 통신 장치, 컴퓨터 장치, 휴대용 멀티미디어 장치, 휴대용 의료 기기, 카메라, 웨어러블 장치, 가전 장치를 포함한다.

**【0128】** 본 발명의 실시예에 따른 방법은 컴퓨터 프로그램 제품에 포함되어 제공될 수 있다. 컴퓨터 프로그램 제품은 상품으로서 판매자 및 구매자 간에 거래될 수 있다. 컴퓨터 프로그램 제품은 기기로 읽을 수 있는 기록매체의 형태로, 또는 어플리케이션 스토어를 통해 온라인으로 배포될 수 있다. 온라인 배포의 경우에, 컴퓨터 프로그램 제품의 적어도 일부는 제조사의 서버, 어플리케이션 스토어의 서버, 또는 중계 서버의 메모리와 같은 저장 매체에 적어도 일시 저장되거나,

임시적으로 생성될 수 있다.

【0129】 본 발명의 실시예에 따른 구성요소, 예컨대 모듈 또는 프로그램 각각은 단수 또는 복수의 서브 구성요소로 구성될 수 있으며, 이러한 서브 구성요소들 중 일부 서브 구성요소가 생략되거나, 또는 다른 서브 구성요소가 더 포함될 수 있다. 일부 구성요소들(모듈 또는 프로그램)은 하나의 개체로 통합되어, 통합되기 이전의 각각의 해당 구성요소에 의해 수행되는 기능을 동일 또는 유사하게 수행할 수 있다. 본 발명의 실시예에 따른 모듈, 프로그램 또는 다른 구성요소에 의해 수행되는 동작들은 순차적, 병렬적, 반복적 또는 휴리스틱하게 실행되거나, 적어도 일부 동작이 다른 순서로 실행되거나, 생략되거나, 또는 다른 동작이 추가될 수 있다.

【0130】 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 구성요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

【0131】 본 발명의 범위는 후술하는 청구범위에 의하여 나타내어지며, 청구 범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

## 【부호의 설명】

【0133】 100: 인공지능 약물 반응성 예측 시스템

110: 유전자 데이터 입력부

120: 유전자 데이터 전처리부

130: 약물 반응성 예측 인공지능 모델

140: 약물 반응성 출력부

## 【청구범위】

### 【청구항 1】

DDPM(Denoising Diffusion Probabilistic Models) 및 그래프 주의 네트워크 기반의 인공지능 약물 반응성 예측 시스템에 있어서, 환자의 유전자 데이터를 수집하는 유전자 데이터 입력부; 타겟 유전자와 생물학적 경로 유전자 간의 근접성을 계산하여 중요한 유전자를 선별하는 유전자 데이터 전처리부; 전처리된 유전자 데이터를 입력받아 DDPM을 통해 원본 데이터와 유사한 증강 데이터를 생성하고, 그래프 주의 네트워크를 통해 약물 반응성을 예측하는 약물 반응성 예측 인공지능 모델; 및 예측된 약물 반응성을 출력하는 약물 반응성 출력부를 포함하는 것을 특징으로 하는 시스템.

### 【청구항 2】

제1항에 있어서, 상기 유전자 데이터 입력부는 암세포주 정보 데이터, GE, 암세포주와 약물 화합물 간의 반응성 정보 데이터를 포함하는 환자 유전자 데이터를 포함하는 것을 특징으로 하는 시스템.

### 【청구항 3】

제1항에 있어서,

상기 유전자 데이터 전처리부는 GDSC(Genomics of Drug Sensitivity in Cancer), TCGA(The Cancer Genome Atlas) 및 PDX(Patient-Derived Xenograft) 중 적어도 하나의 데이터베이스에 접속하여 데이터를 입수하는 것을 특징으로 하는 시스템.

### 【청구항 4】

제1항에 있어서,

상기 유전자 데이터 전처리부는 생물학적 경로를 분석하여 중요한 유전자를 선별하고, 그래프 오토인코더(GAE)를 통해 생물학적 관계를 반영하는 잠재 공간을 생성하는 것을 특징으로 하는 시스템.

### 【청구항 5】

제1항에 있어서,

상기 약물 반응성 예측 인공지능 모델은 상기 DDPM을 이용하여 증강된 데이터의 일반화를 수행하는 것을 특징으로 하는 시스템.

## 【청구항 6】

제2항에 있어서,

상기 약물 반응성 예측 인공지능 모델은 그래프 주의 네트워크를 통해 타겟 단백질과 경로 유전자 간의 생물학적 경로를 기반으로 약물 반응성을 예측하며, 상기 그래프 주의 네트워크는 약물 타겟 유전자와 관련된 생물학적 경로를 기반으로 그래프 구조를 형성하고, 각 유전자가 그래프의 노드로 표현되며, 노드 간 연결 정보를 통해 유전자 간 관계를 반영하여 그래프 학습이 수행되며, 각 생물학적 경로에 대해 독립적인 그래프 주의 연산을 수행한 후, 각 노드 정보의 결합을 통해 약물 반응성에 대한 예측 결과를 생성하는 것을 특징으로 하는 시스템.

## 【청구항 7】

약물 반응성 예측 시스템이 약물 반응성 예측을 위한 방법에 있어서,

환자의 유전자 데이터를 수집하는 단계;

타겟 유전자와 경로 유전자 간의 근접성을 계산하여 중요한 유전자를 선별하는 단계;

선별된 유전자 데이터를 DDPM(Denoising Diffusion Probabilistic Models)을 이용해 증강 데이터를 생성하는 단계;

증강된 데이터를 그래프 주의 네트워크를 통해 약물 반응성을 예측하는 단계; 및

상기 예측된 약물 반응성을 출력하는 단계

를 포함하는 것을 특징으로 하는 약물 반응성 예측 방법.

### 【청구항 8】

제7항에 있어서,

상기 환자의 유전자 데이터를 수집하는 단계는 암세포주 정보 데이터, GE, 암세포주와 약물 화합물 간의 반응성 정보 데이터를 포함하는 유전자 데이터를 수집하는 것을 특징으로 하는 방법.

### 【청구항 9】

제7항에 있어서,

상기 유전자 데이터 전처리 단계는 GDSC(Genomics of Drug Sensitivity in Cancer), TCGA(The Cancer Genome Atlas) 및 PDX(Patient-Derived Xenograft) 중 적어도 하나이 데이터베이스와 접속하여 데이터를 입수하는 것을 특징으로 하는 방법.

### 【청구항 10】

제7항에 있어서,

상기 유전자 데이터 전처리 단계는 생물학적 경로를 분석하여 중요한 유전자

를 선별하고, 그래프 오토인코더(GAE)를 통해 생물학적 관계를 반영하는 잠재 공간을 생성하는 것을 특징으로 하는 방법.

### 【청구항 11】

제7항에 있어서,  
약물 반응성을 예측하는 단계는 그래프 주의 네트워크를 통해 타겟 단백질과 경로 유전자 간의 생물학적 경로를 기반으로 약물 반응성을 예측하는 것을 특징으로 하는 방법.

### 【청구항 12】

제11항에 있어서,  
상기 그래프 주의 네트워크는 약물 타겟 유전자와 관련된 생물학적 경로를 기반으로 그래프 구조를 형성하고, 각 유전자가 그래프의 노드로 표현되며, 노드 간 연결 정보를 통해 유전자 간 관계를 반영하여 그래프 학습이 수행되며, 각 생물학적 경로에 대해 독립적인 그래프 주의 연산을 수행한 후, 각 노드 정보의 결합을 통해 약물 반응성에 대한 예측 결과를 생성하는 것을 특징으로 하는 방법.

## 【요약서】

### 【요약】

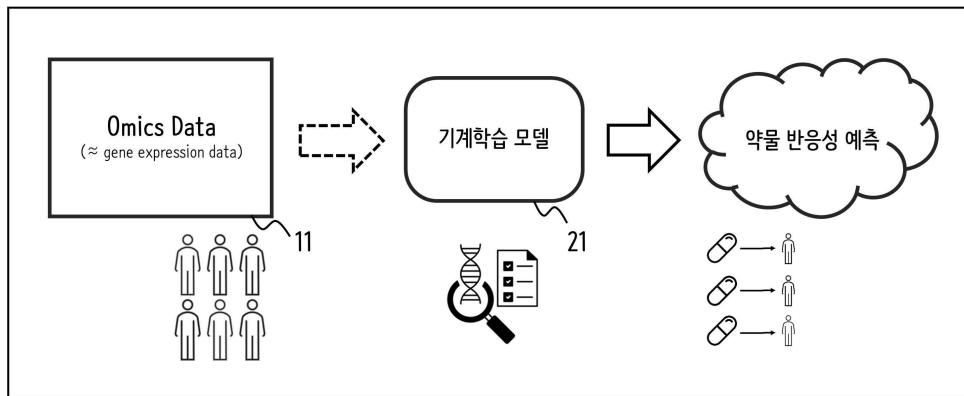
본 발명에 따른 DDPM 및 그래프 주의 네트워크 기반의 인공지능 약물 반응성 예측 시스템은, 환자의 유전자 데이터를 수집하는 유전자 데이터 입력부와 타겟 유전자와 생물학적 경로 유전자 간의 근접성을 계산하여 중요한 유전자를 선별하는 유전자 데이터 전처리부와 전처리된 유전자 데이터를 입력받아 DDPM을 통해 원본 데이터와 유사한 증강 데이터를 생성하고, 그래프 주의 네트워크를 통해 약물 반응성을 예측하는 약물 반응성 예측 인공지능 모델 및 예측된 약물 반응성을 출력하는 약물 반응성 출력부를 포함한다.

### 【대표도】

도 2

## 【도면】

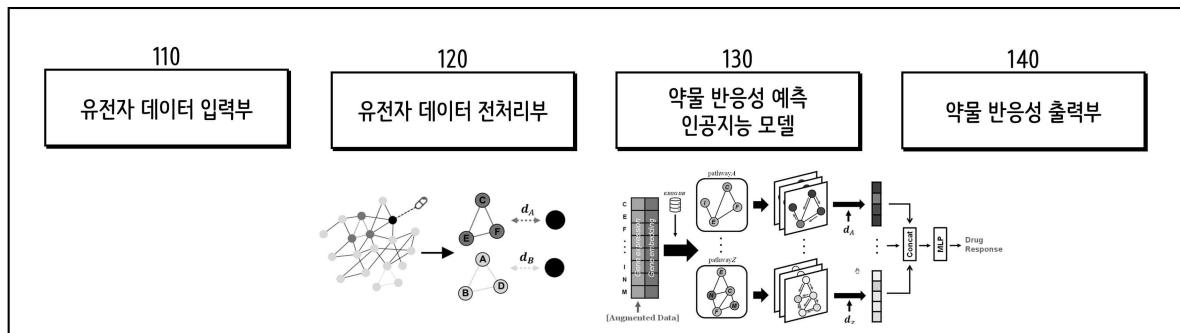
## 【도 1】



101

## 【도 2】

100 DDPM 및 그래프 주의 네트워크 기반의 인공지능 약물 반응성 예측 시스템



## 【도 3】

S110

유전자 데이터베이스로부터 환자 유전자 데이터를 입수하는 단계

S120

유전자의 생물학적 경로를 식별하여  
중요 경로 유전자 데이터를 선별하는 단계

S130

선별된 유전자 데이터를 입력으로  
DDPM 생성 모델을 실행하는 단계

S140

약물 반응성 예측 인공지능 모델에 의해  
약물 반응성을 예측하고 예측 결과를 제공하는 단계

## 【도 4】

S121

입력된 유전자 데이터로부터  
유전자 집합 구성

S122

타겟 유전자와 관련된  
생물학적 경로 식별

S123

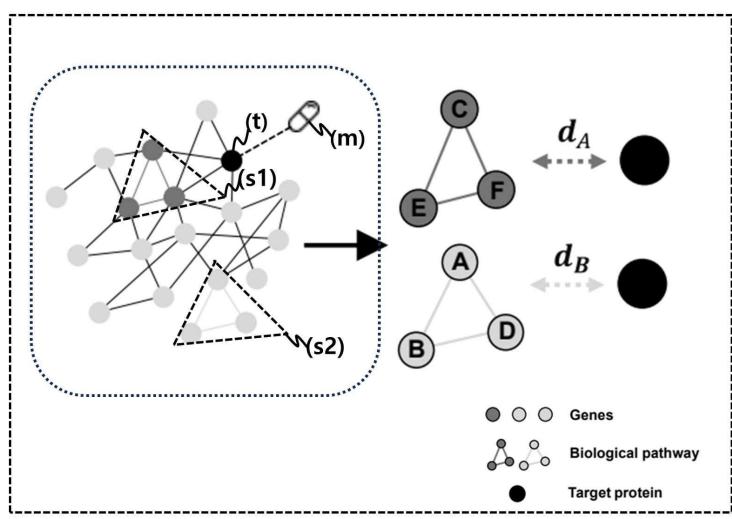
Z-점수를 산출하여  
경로 유전자의 기준 분포 생성

S124

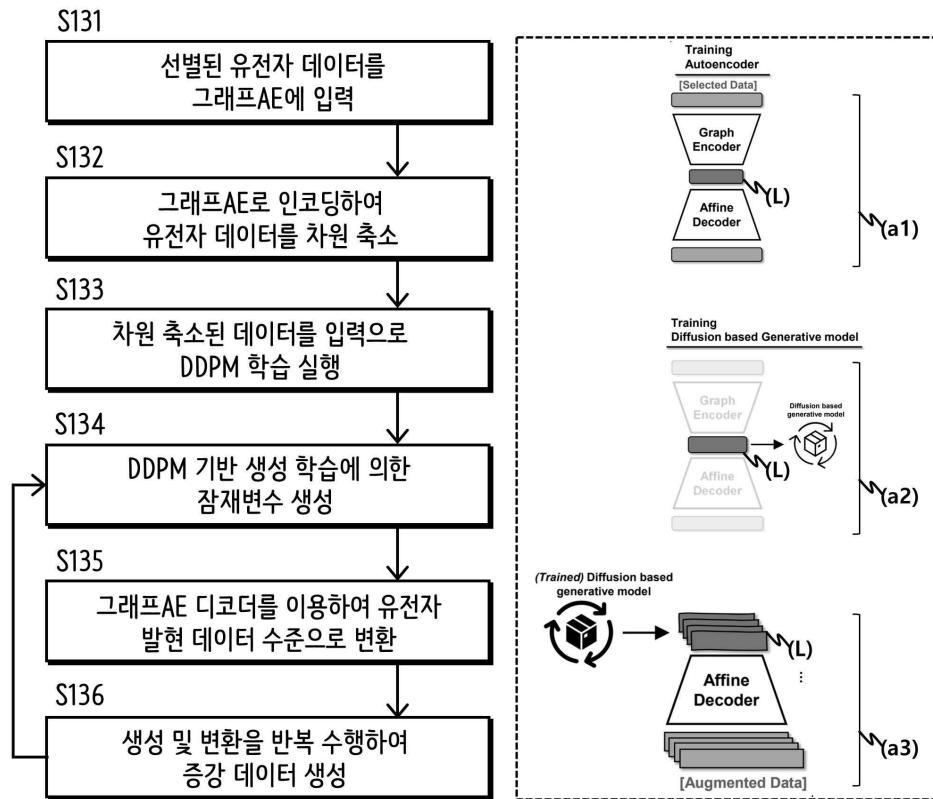
기준 분포를 적용하여  
통계적 관련성 지표 도출

S125

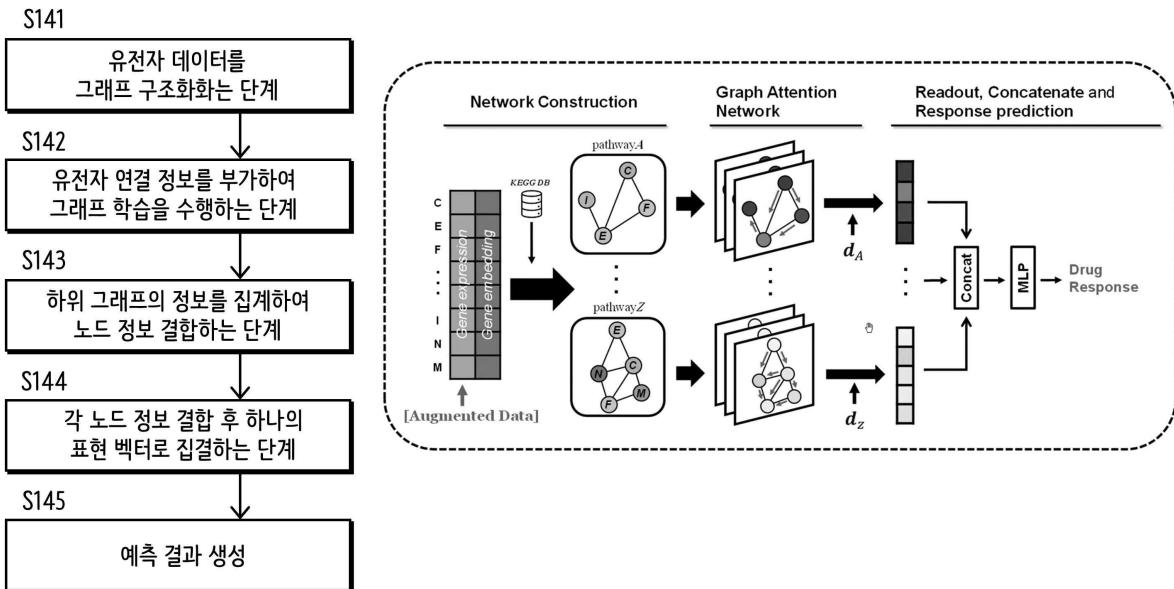
경로 유전자 선별



## 【도 5】



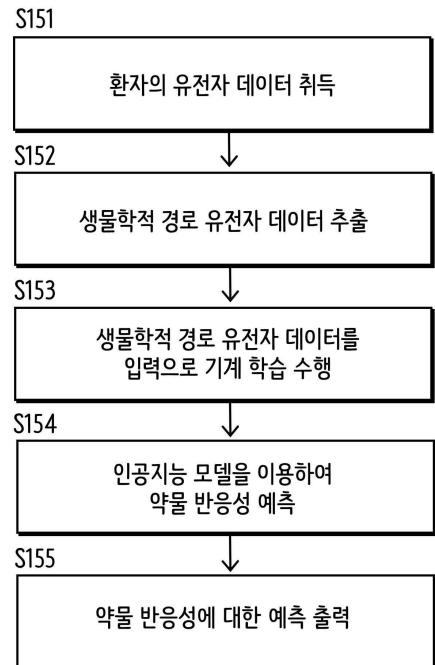
## 【도 6】



## 【도 7】

Model	PDX Dataset			TCGA Dataset			Average
	<i>Paclitaxel</i>	<i>Cetuximab</i>	<i>Erlotinib</i>	<i>Docetaxel</i>	<i>Cisplatin</i>	<i>Gemcitabine</i>	
NX	0.24	0.53	0.28	0.39	0.40	0.58	0.40
FX	0.68	0.43	0.37	0.69	0.44	0.65	0.54
GX	0.52	0.58	0.67	0.59	0.62	0.53	0.58
AX	0.44	0.42	0.33	0.50	0.46	0.50	0.44
AXX	0.64	0.43	0.65	0.64	0.68	0.57	0.60
AXXX	0.46	0.17	0.17	0.42	0.45	0.53	0.36
SXXX	0.49	0.41	0.67	0.53	0.47	0.47	0.50
TXXX	0.51	0.51	0.58	0.50	0.50	0.50	0.51
MXXX	0.74	0.53	0.63	0.58	0.66	0.65	0.63
SXX	0.64	0.55	0.76	0.64	0.73	0.61	0.65
DX	0.74	0.71	0.85	0.78	0.68	0.53	0.71
(r) DiffGA (No-Aug)	0.70	0.67	0.80	0.72	<b>0.75</b>	<b>0.66</b>	0.71
DiffGA	<b>0.78</b>	<b>0.82</b>	0.83	<b>0.83</b>	<b>0.81</b>	<b>0.72</b>	<b>0.78</b>

## 【도 8】



## 【도 9】

