

발간등록번호

11-1471057-100001-01

국민 안전이 기준입니다
YOUR SAFETY IS OUR STANDARD

인공지능(AI) 활용 의약품 개발 사례집

2024. 12.



식품의약품안전처

식품의약품안전평가원

의약품심사부 순환신경계약품과

안 내 사 항

인공지능(AI)/기계학습(ML) 기술 도입으로 의약품 개발 기간 단축 및 비용 절감, 약물 재창출 등 의약품 개발 패러다임이 빠르게 전환되고 있습니다. 이에 발맞추어 식품의약품안전처에서는 인공지능(AI)을 활용한 의약품 개발 시 참고 자료로 활용할 수 있도록 『의약품 개발 시 인공지능(AI) 활용 안내서』(‘24.6)를 마련한 바 있습니다. 이 안내서는 의약품 개발 단계별 인공지능(AI)/기계학습(ML) 활용 범위 및 인공지능(AI) 기술의 규제적용을 위하여 기술적·윤리적 측면에서 고려해야 할 사항을 담고 있습니다.

본 『인공지능(AI) 활용 의약품 개발 사례집』은 1. 서론, 2. 국외 규제기관의 인공지능(AI)/기계학습(ML) 활용 동향, 3. 인공지능(AI)/기계학습(ML) 활용 의약품 개발 사례, 4. 거버넌스 및 신뢰 기반 인공지능(AI)에 대한 업계 의견으로 구성되어 있습니다. 의약품 개발 단계별 인공지능(AI) 활용 현황·사례 및 인공지능(AI) 기술 규제적용 시 고려해야 하는 기술적·윤리적 측면에 대해 현재 시점에서의 제약업계가 바라보는 시각을 담고 있어 의약품 개발 관련 기술혁신주체(기업·대학 및 연구기관) 및 허가·심사 담당자에게 의약품 개발 과정에서 인공지능(AI) 활용 범위 및 규제 측면에서 고려해야 할 사항에 대한 이해를 높이고, 실질적인 도움이 될 수 있을 것으로 기대합니다.

본 사례집은 대외적으로 법적 효력을 가지는 것이 아니므로 본문의 기술방식(‘~하여야 한다’ 등)에도 불구하고 참고로만 활용하시기 바랍니다. 또한, 2024년 12월 현재의 과학적·기술적 사실 및 유효한 법규를 토대로 작성되었으므로 이후 최신 개정 법규 내용 및 구체적인 사실관계 등에 따라 달리 적용될 수 있음을 알려드립니다.

※ 본 안내서에 대한 의견이나 문의사항이 있을 경우 식품의약품안전처 식품의약품안전평가원 의약품심사부 순환신경계약품과에 문의하시기 바랍니다.

전화 : 043-719-3002, 3006

팩스 : 043-719-3000

목 차

1. 서론	1
2. 국외 규제기관의 인공지능(AI)/기계학습(ML) 활용 동향	3
2.1. 미국 FDA	3
2.2. 유럽 EMA	5
3. 인공지능(AI)/기계학습(ML) 활용 의약품 개발 사례	7
3.1. 후보물질 발굴	9
3.2. 비임상시험	10
3.3. 임상시험	13
4. 거버넌스 및 신뢰 기반 인공지능(AI)에 대한 업계 의견	23
4.1. 거버넌스, 책임 및 투명성	23
4.2. 데이터 관련 고려사항	24
4.3. 모델 관련 고려사항	26

1

서론

인공지능(Artificial Intelligence, AI)이 산업계 전반에서 혁신적인 변화를 이끌며, 특히 효율성과 생산성을 극대화하는 데 기여하고 있다. 인공지능(AI)은 산업·사회 구조의 광범위한 변화를 일으키는 혁신 기술(innovative technologies)로서, 이러한 기술적 혁신에 대응하기 위해 주요 국외 규제기관들은 인공지능(AI) 기술을 안전하고 효과적으로 활용하기 위한 정책 및 법률 마련에 박차를 가하고 있다. 최근 유럽연합(EU)에서는 인공지능(AI) 관련 기술 허용 범위와 규제 대상 등을 규정한 세계 최초 인공지능 규제법인 「AI 법(AI act)」을 승인하였으며(‘24.5., ‘26년 시행), 미국에서는 안전하고 보안이 보장되며 신뢰할 수 있는 인공지능(AI) 개발 및 사용에 대한 AI 행정 명령(‘23.10.30)을 내리는 등 세계 주요국들은 합리적인 규제 프레임에 마련하기 위해 노력하고 있다. 우리나라도 인공지능(AI) 기술 산업의 건전한 발전을 지원하고 인공지능(AI) 사회의 신뢰 기반 조성에 필요한 기본적인 사항을 규정하기 위하여 AI 기본법(가칭)으로 불리는 법률안이 발의되어 논의 과정 중에 있다. 참고로 식품의약품안전처(이하 ‘식약처’)는 디지털 특성에 기반한 디지털 의료제품의 안전성과 유효성을 평가할 수 있도록 안전관리 체계를 별도로 마련하여 디지털 의료제품의 발전을 지원하기 위한 세계 첫 「디지털의료제품법」 시행을 앞두고 있다(‘25.1.).

의료시장에서 인공지능(AI) 기술 활용 사례가 증가하고 있으며, 제약산업 분야에서도 인공지능(AI) 기술이 전통적인 의약품 개발의 한계를 극복할 수 있는 핵심 도구(tool)로 주목받고 있다. 일반적으로 신약 개발은 평균적으로 10년 이상의 시간이 소요되고 성공 가능성이 약 10% 정도로 매우 낮다. 인공지능(AI)은 방대한 데이터를 분석해 후보물질 탐색·발굴, 흡수·분포·대사·배설·독성(ADMET) 예측, 임상시험 설계 등 의약품 개발의 다양한 단계에 활용되어 개발 소요 기간 및 비용을 획기적으로 감축할 것으로 예상된다.

이러한 흐름에 발맞추어 미국 FDA(Food and Drug Administration, FDA)는 인공지능(AI)/기계학습(Machine Learning, ML) 사용 의약품 개발에 대한 보고서(‘23.5) 초안을 발표한 바 있고, 유럽의약품청(European Medicines Agency, EMA)에서는 의약품 전주기 단계에서 인공지능(AI)/기계학습(ML) 사용원칙에 대한 의견서(초안 ‘23.7., 최종 ‘24.9.)를 발표하는 등 의약품 개발 시 인공지능 활용에 대한 규제기관의 현재 입장과 관점을 제시하고 있다. 식약처도 의약품 개발 시 인공지능(AI)/기계학습(ML) 활용 분야 및 데이터·인공지능(AI) 기술 관련 규제적용을 위하여 기술적, 윤리적 측면에 있어 고려해야 할 사항을 제시한 「의약품 개발 시 인공지능(AI) 활용 안내서」(‘24.6)를 마련하였다. 이 안내서는 의약품 분야에서 논의가 이루어지고 있는 현재까지의 국내·외 인공지능(AI) 규제 상황 등을 검토하여 마련한 것으로 확정된 원칙을 제시하는 것은 아니며, 인공지능(AI) 기술을 적용한 의약품 개발 시 참고자료로 활용할 수 있도록 구성하였다.

본 사례집은 인공지능(AI)/기계학습(ML) 활용 동향 및 국내·외 인공지능(AI)/기계학습(ML) 활용 의약품 개발 사례 등을 제공하기 위해 의약품심사소통단(CHORUS)* 허가·심사 지원 분과의 ‘인공지능(AI) 활용 신약개발’ 소분과**와 함께 논의하여 마련하였다. 국외 규제당국 및 문헌자료 기반 인공지능(AI)/기계학습(ML) 적용 사례 및 국내 제약기업의 인공지능(AI) 활용 사례를 제공하고, 인공지능(AI) 기술 규제적용 시 고려해야 하는 기술적·윤리적 측면에 대한 현재 시점에서의 제약업계가 바라보는 시각을 담고 있다.

* 의약품심사소통단(CHORUS): 국내 제약산업 성장 지원을 위한 식약처와 업계 간 의약품 심사분야 민·관 소통 채널

** ‘인공지능(AI) 활용 신약개발’ 소분과 구성: 인공지능(AI) 활용 의약품 개발업체(7개사), 한국제약바이오협회 (AI신약융합연구원), 한국의약품수출입협회, 식약처

본 사례집이 의약품 개발 관련 기술혁신주체(기업·대학 및 연구기관) 및 허가·심사 담당자에게 의약품 개발 과정에서 인공지능(AI)/기계학습(ML) 활용 범위 및 규제 측면에서 고려해야 할 사항에 대한 이해를 높이고, 실질적인 도움이 될 수 있을 것으로 기대한다.

2

국외 규제기관의 인공지능(AI)/기계학습(ML) 활용 동향

의약품 개발 분야에서 인공지능(AI)/기계학습(ML)의 활용이 확대됨에 따라, 인공지능(AI)/기계학습(ML)을 안전하고 책임감 있게 활용하기 위한 규제의 중요성이 더욱 커지고 있다. 이에 따라 주요 국외 규제기관들은 인공지능(AI)/기계학습(ML)의 안전하고 효과적인 사용을 지원하기 위해 기술적, 윤리적 관점에서의 고려사항 등을 제시한 문서를 마련하였으며, 이를 포함한 주요 국외(미국, 유럽) 규제기관의 인공지능(AI)/기계학습(ML) 활용 동향은 다음과 같다.

2.1. 미국 FDA

미국 FDA 내부 데이터베이스에서 '16~'21년도까지 제출된 자료를 '인공지능(AI)' 또는 '기계학습(ML)'으로 검색하여 분석한 결과, '16 및 '17년도에는 각 1건에 불과했던 사례가 '21년도에는 132건으로 비약적으로 증가했으며, 의약품 개발 단계별로 구분하였을 때 임상 개발에 활용된 사례가 140건, 비임상 8건, 시판 후 조사 6건, 후보물질 발굴/약물재창출 4건이었다.¹⁾ 인공지능(AI)/기계학습(ML)은 다양한 치료 분야에서 활용되고 있었으며, 항암제(27%)가 가장 높은 비중을 차지했고, 정신계, 소화기계, 신경계 순이었다(그림 1).

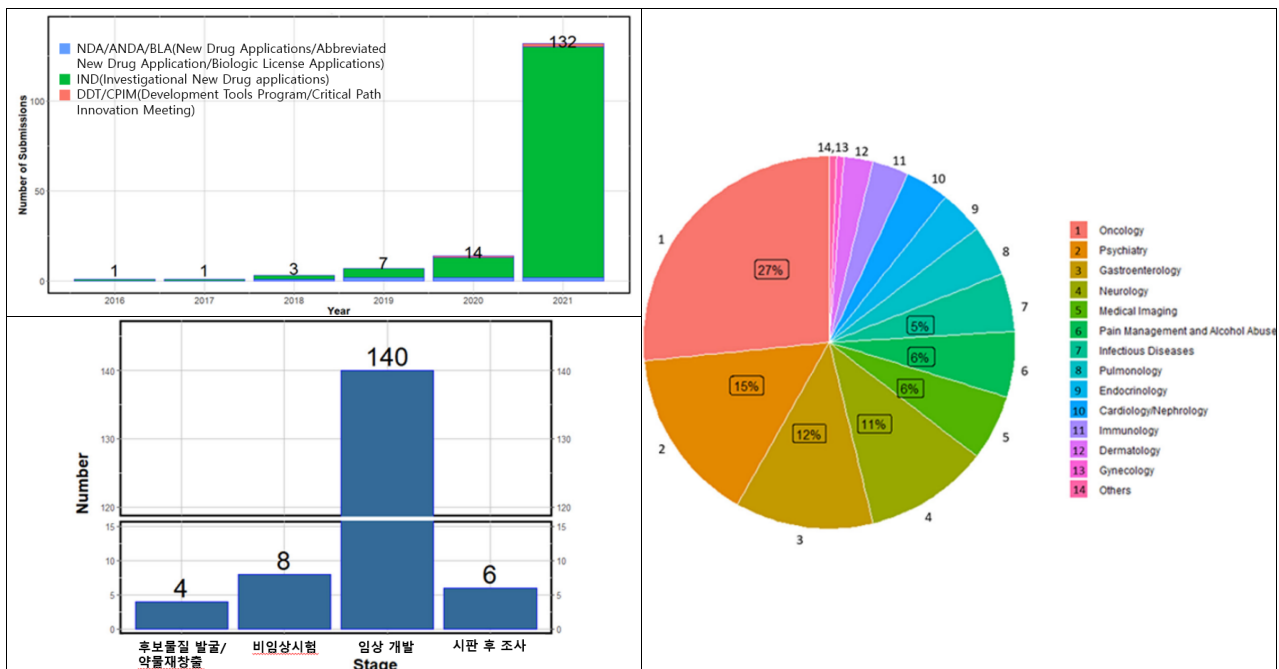


그림 1. 미국 FDA로 제출된 인공지능(AI)/기계학습(ML) 활용 의약품 개발 현황¹⁾

1) Liu Q, et al. Landscape Analysis of the Application of Artificial Intelligence and Machine Learning in Regulatory Submissions for Drug Development From 2016 to 2021. Clin Pharmacol Ther. 2023;113(4):771-774.

미국 FDA의 CDER(Center for Drug Evaluation and Research)는 의약품 전주기에서 인공지능(AI)/기계학습(ML) 활용에 대응하고자 인공지능 운영 위원회(AI steering committee, AISC)를 조직화하였으며('20), CDER 내부의 초기 인공지능(AI) 관련 활동을 주도하며 정책, 규제 및 기술적 접근 방식에 대한 협업과 조정을 담당하였다. 이후 AI 행정 명령('23.10.30) 등 외부 환경이 변화함에 따라, '24년 CDER 인공지능 위원회(AI council)가 조직되어 인공지능 운영 위원회(AISC)를 포함한 다양한 AI 관련 그룹(AI Policy Working Group, CDER AI Community of Practice)의 작업을 통합하고 체계화된 방식으로 이어가게 되었다. '인공지능 위원회'는 의사결정 기구로서 의약품의 안전성·유효성 및 품질 평가와 관련된 인공지능(AI) 고려사항에 대한 일관성을 높이고 CDER 내부 인공지능(AI)의 적절한 활용 확대를 위한 중추적 역할을 담당한다.²⁾

미국 FDA는 비임상·임상시험 분야에서 인공지능(AI)/기계학습(ML)을 활용한 다양한 경험을 축적하고 있으며 주요 사례는 다음과 같다.

활용 분야	주요 내용
신약 개발 도구 (Drug Development Tool, DDT)	신약에 대한 혁신적인 과학기술 접근법(Innovative Science and Technology Approaches for New Drugs, ISTAND) 파일럿 프로그램에서 인공지능(AI) 기반 디지털 헬스기술 프로젝트* 채택('23.11.) ³⁾ * 기계학습(ML) 모델을 활용하여 우울증과 불안의 여러 행동적, 생리적 지표를 기반으로 우울증 및 불안의 중증도를 자동으로 측정하는 도구
의약품 투여가 적절한 환자군 확인	인공지능(AI)/기계학습(ML) 기반 의약품 투여가 적절한 환자군 확인(3.3항 참조) ⁴⁾ - (품목명) 키너렛주(KINERET® injection) - (활용내용) 수용성 유로키나아제 플라스미노겐 활성제 수용체(suPAR) ≥6ng/mL인 환자를 예측할 수 있는 점수 측정 도구 개발
허가사항 검토 지원	FDA 산하 국립독성연구센터(NCTR)에서 대규모 언어모델(Large Language Model, LLM) 기반 허가사항 검토 지원 프로그램 'askFDALabel' 개발('24.5.) ⁵⁾
비임상시험	FDA 산하 국립독성연구센터(NCTR)에서 인공지능(AI) 기술을 활용하여 새로운 평가도구 및 규제과학 지원을 위한 세부 이니셔티브 구성('22~) ⁶⁾ - (AnimalGAN) 생성형 인공지능(GenAI)을 활용하여 디지털 트윈(digital twin)으로 가상 동물 모델을 개발하고, 3R 원칙(동물 실험의 대체, 감소, 개선)을 지원 - (SafetAI) 독성학적 평가변수를 분석하기 위한 인공지능 모델 개발을 목적으로 하며, 간독성, 발암성, 돌연변이시험, 신독성, 심장독성에 중점을 둠 - (BERTox) 자연어 처리(Natural Language Processing, NLP) 및 대규모 언어 모델(LLM)을 활용하여 FDA 내부 문서와 공개된 문헌을 분석하고, 규제과학 및 검토 과정의 효율성과 정확성을 향상시키는 것을 목표로 함 - (PathologAI) 동물실험에서 얻은 조직병리학적 데이터를 분석하기 위한 효과적이고 정확한 프레임워크 개발을 목표로 함 - (TranslAI) 생성형 인공지능(GenAI) 모델을 개발하여 기관계, 생체 내·외 외삽(in vitro-to-in vivo extrapolation, IVIVE), 유전체 기술과 같은 영역에서 연구 결과를 효과적으로 중개할 수 있도록 지원

2) <https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/artificial-intelligence-drug-development> ('24.12.16. 기준)

또한 미국 FDA는 인공지능(AI)/기계학습(ML) 활용에 대한 의견을 수집하기 위해 의약품 개발 분야에서 인공지능(AI)/기계학습(ML) 활용 가능 분야 및 활용 시 고려사항을 제시한 ‘의약품 및 생물학의약품 개발에서 인공지능(AI)/기계학습(ML) 활용-토의문서(Using Artificial Intelligence & Machine Learning in the Development of Drug & Biological Products, ‘23.5)’를 발표하였으며,⁷⁾ 수집된 의견 및 CDER의 경험 등을 반영하여 ‘의약품 및 생물학의약품에 대한 규제적 의사결정을 지원하기 위해 인공지능(AI) 활용 시 고려사항에 대한 가이드라인(Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drugs and Biological Product, ‘24)’을 마련할 예정이다.⁸⁾

2.2. 유럽 EMA

유럽 EMA와 유럽 의약품 안전관리기구(Heads of Medicines Agency, HMA)로 구성된 빅데이터 조정 그룹(Big Data Steering Group)에서 ‘AI Workplan 2023-2028’을 발표하였으며(‘23.11.), 본 계획에서는 4가지 주요 영역에서 의약품 규제 과정에 인공지능(AI)을 통합하기 위한 전략을 제시하였다.⁹⁾

구분	주요 내용
지침, 정책 및 제품화 지원 (Guidance, Policy and Product support)	· 의약품 전주기에서 인공지능(AI) 활용에 대한 지침 제공 * 주요 추진계획 : (‘24~) 의약품 전주기에서 AI 관련 가이드라인 개발 착수 등
도구 및 기술 (Tools & Technologies)	· 인공지능(AI) 도구 사용을 위한 프레임워크 제공 * 주요 추진계획 : (‘24~) 내부 직원 대상 AI 도구 사용에 대한 지침 발표 등
협력 및 교육 (Collaboration and change management)	· 인공지능(AI) 기술 사용 능력 개발 * 주요 추진계획 : (‘24~) 연 1회 공개 워크숍 개최 등
실험 (Experimentation)	· 실험을 통해 시스템에 대한 불확실성을 감소시키고 구조적이고 조정된 접근 방식 보장 * 주요 추진계획 : (‘24~) 네트워크 연구 우선순위 로드맵 발표 등

3) <https://www.fda.gov/drugs/drug-safety-and-availability/fdas-istand-pilot-program-accepts-submission-first-artificial-intelligence-based-and-digital-health>(‘24.12.16. 기준)

4) <https://www.fda.gov/drugs/spotlight-cder-science/using-machine-learning-identify-suitable-patient-population-anakinra-treatment-covid-19-under>(‘24.12.16. 기준)

5) <https://www.fda.gov/about-fda/nctr-publications/nctr-research-highlights>(‘24.12.16. 기준)

6) <https://www.fda.gov/about-fda/nctr-research-focus-areas/artificial-intelligence>(‘24.12.16. 기준)

7) <https://www.fda.gov/media/167973/download>(‘24.12.16. 기준)

8) CDER Guidance Agenda(<https://www.fda.gov/media/134778/download?attachment>, ‘24.12.16. 기준)

9) <https://www.ema.europa.eu/en/about-us/how-we-work/big-data/artificial-intelligence>(‘24.12.16. 기준)

유럽 EMA에서는 의약품 전주기의 다양한 단계에서 인공지능(AI)/기계학습(ML)을 안전하고 효과적으로 사용하기 위한 고려사항 등을 제시한 ‘의약품 전주기 단계에서 인공지능(AI)/기계학습(ML) 사용 원칙에 대한 의견서(Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle, ‘24.9)’를 발표하였다.¹⁰⁾ 또한 유럽 규제 당국을 위한 인공지능(AI) 기반 지식 탐색 도구인 ‘Scientific Explorer’를 도입하여(‘24.3) 규제과학 정보를 손쉽게, 집중적이며 정확하게 검색할 수 있도록 지원하였으며,¹¹⁾ 직원들이 대규모 언어모델(LLM)을 업무에 활용할 수 있도록 지원하는 원칙을 마련하여 대규모 언어모델(LLM)을 안전하고 효과적으로 사용할 수 있도록 지원하였다(‘24.8).¹²⁾

아울러, ‘AIM-NASH(Artificial Intelligence-Based Measurement of Non-alcoholic Steatohepatitis Histology)’라는 인공지능(AI) 기반 병리학적 평가도구에 대해 의견조회(‘24.12.6.~‘25.1.24.) 중이다.¹³⁾ AIM-NASH는 대사 기능 장애 관련 지방간염(metabolic dysfunction associated steatohepatitis, MASH) 임상시험에서 간 생검의 조직 병리학적 특징을 분석하고, 질병 활동성을 평가하는 데 보조 도구로 활용될 수 있으며, 비알코올성 지방간질환 활성도 점수(Non-alcoholic Fatty Liver Disease Activity Score, NAS)의 구성 요소(지방증(steatosis), 간세포 팽창(hepatocellular ballooning), 간소엽의 염증세포침윤(lobular inflammation))와 섬유화 단계를 산출한다. 이를 통해 임상 2상 및 3상 시험에서 초기 평가와 추적 관찰 시 활용할 수 있도록 설계된 도구이다.

10) <https://www.ema.europa.eu/en/use-artificial-intelligence-ai-medicinal-product-lifecycle>(‘24.12.16. 기준)

11) https://www.ema.europa.eu/system/files/documents/other/scientific-explorer-frequently-asked-questions_en.pdf (‘24.12.16 기준)

12) https://www.ema.europa.eu/en/documents/other/guiding-principles-use-large-language-models-regulatory-science-medicines-regulatory-activities_en.pdf(‘24.12.16. 기준)

13) https://www.ema.europa.eu/en/documents/other/draft-qualification-opinion-artificial-intelligence-based-measurement-non-alcoholic-steatohepatitis-histology-liver-biopsies-determine-disease-activity-nash-mash-clinical-trials_en.pdf (‘24.12.16. 기준)

3

인공지능(AI)/기계학습(ML) 활용 의약품 개발 사례

인공지능(AI)은 학습, 추론, 판단, 언어의 이해 등 인간이 가진 지적 능력을 전자적 방법으로 구현하는 기술이며, 기계학습(ML)은 기계가 데이터를 기반으로 자료의 패턴을 찾아 인공지능을 수행할 수 있도록 하는 것으로, 과거의 규칙 기반 알고리즘과 대비되며 대량의 자료와 결과를 투입하면 컴퓨터가 투입된 자료와 결과를 연결하는 논리를 만들어낸다는 개념이다.

인공지능(AI)/기계학습(ML)은 의약품 개발 시 후보물질 발굴, 비임상시험, 임상시험에 활용되어 시간·비용적 측면에서 효율성을 개선할 수 있다. 국내 제약업계의 인공지능(AI) 모델 활용 현황을 조사한 결과(한국제약바이오협회 AI신약융합연구원 제공), 대부분 후보물질 발굴 목적으로 인공지능(AI) 모델을 활용하고 있었으며, 의약품 설계 등 각 분야별 사용되는 기술은 다음과 같다.

단계	분야	기술
후보물질 발굴	의약품 설계	Protein Folding
		Protein Structure Prediction
		Protein-Protein Interaction
		Target-Protein Binding Prediction
		Target-Ligand Binding Prediction
		Target-Peptide Binding Prediction
		Similarity Search
		Fragment Based Drug Discovery
		Scaffold Hopping
		Generation Model
		Natural Language Processing
		Retrosynthesis
	의약품 물성, 활성 및 독성 예측	Graph Neural Network
		Similarity Search
		Target-Ligand Binding Prediction
		QSAR
		Transcriptomics Analysis
		Generation Model
		Self-Supervised Learning

단계	분야	기술
	약물 특성 및 약물작용기전 예측	Network modeling
		Cutting-edge Model
		Natural Language Processing
		Multimomics analysis
		Proteomics Analysis
		Transcriptomics Analysis
		Explainable AI
	약물 표적 및 생체지표(biomarker) 발굴	Multimomics Analysis
		Proteomics Analysis
		Genomics Analysis
		scRNAAnalysis
		Cell Image Data Analysis
		Network Modeling
		Image Analysis
	표적-약물 결합 예측	Quantum Chemical Calculation
		Target-Ligand Binding Prediction
		Fragment Based Drug Discovery
		Pharmacokinetics Data Analysis
		Graph Neural Network
		Conformational Analysis
		Quantum mechanics
		Molecular Dynamics
		Pharmacophore prediction
후보물질 발굴 (약물재창출 등) 및 비임상 · 임상	약물재창출	Combinatorial Optimization
		Transcriptomics Analysis
		Similarity Search
	임상시험 데이터 분석 약물 복합제 가능성 예측	Cell Image Data Analysis
		Natural Language Processing
		Network Modeling

3.1. 후보물질 발굴

인공지능(AI)/기계학습(ML)은 후보물질 발굴 단계에서 약물 표적 확인, 화합물 스크리닝 및 설계 등에 활용될 수 있다. 국내 제약업계에서는 인공지능(AI)/기계학습(ML)을 대부분 후보물질 발굴 단계에서 활용하고 있었으며, 의약품심사소통단(CHORUS) 허가·심사 지원 분과 ‘인공지능(AI) 활용 신약개발’ 소분과에 참여한 제약회사 및 AI신약융합연구원 소속 AI 신약개발 전문위원회 (22개사)를 대상으로 조사한 인공지능(AI)/기계학습(ML) 활용 사례는 다음과 같다.

회사명	치료분야	연구단계	AI 적용분야	협업기업명
대웅제약	대사질환	선도물질 최적화	화합물 스크리닝 및 디자인	-
			신규 화합물 생성(GenAI)	(자체 플랫폼)
	항암	선도물질 도출	화합물 디자인(FBDD)* * FBDD(Fragment-based drug discovery): 작은 사이즈의 fragment hit이 주어졌을 때 이를 키우거나 (growing) 여러 개를 연결(linking)하여 drug-like한 화합물을 생성하는 플랫폼	-
	자가면역질환	선도물질 도출		(자체 플랫폼)
	항암	전임상	화합물 스크리닝 및 디자인	XtalPi
	자가면역질환	전임상	화합물 스크리닝 및 디자인	닥터노아바이오텍
동아ST	면역항암	유효물질 도출	화합물 스크리닝 및 디자인	-
	항암	유효물질 도출	화합물 스크리닝 및 디자인	(자체 플랫폼)
	항암	유효물질 도출	화합물 스크리닝 및 디자인	갤럭시
	항암	유효물질 도출	화합물 스크리닝 및 디자인	심플렉스
동화약품	면역질환	선도물질 최적화	유도체 설계 및 디자인	-
	항암제	임상시험	적응증 발굴	(자체 플랫폼)
삼진제약	알러지	선도물질 최적화	화합물 디자인	심플렉스
	항암	선도물질 최적화		온코크로스
	항암	선도물질 최적화	화합물 스크리닝 및 디자인	-
	항암	유효물질 도출	화합물 스크리닝 및 디자인	(자체 플랫폼)
	항암	선도물질 도출	화합물 스크리닝	인세리브로
	항암	선도물질 도출	화합물 스크리닝 및 디자인	아론티어
JW중외제약	대사질환	선도물질 최적화	화합물 스크리닝	심플렉스
	면역질환	임상시험	화합물 스크리닝 및 디자인	-
	희귀질환	선도물질 최적화	화합물 스크리닝 및 디자인	(자체 플랫폼 JWave)
한미약품	대사질환, 항암	선도물질 도출, 최적화 및 활성 예측	단백질/펩타이드 구조 예측	-
			아날로그 활성/안정성/물성 예측	(오픈소스 및 자체 플랫폼 HARP*)
			화합물 도출 및 활성 예측	* HARP: Hanmi AI-driven Research Platform
	항암		화합물 디자인 및 활성 예측	아이젠사이언스

3.2. 비임상시험

인공지능(AI)/기계학습(ML)은 비임상시험 단계에서 흡수·분포·대사·배설·독성(absorption, distribution, metabolism, excretion, toxicity, ADMET) 예측, 작용기전 탐색, 생체 내·외 기전 연구 등에 활용될 수 있다.

의약품 개발 초기 단계에서 흡수·분포·대사·배설·독성(ADMET)은 매우 중요한 요소이며, 신약 개발 R&D 비용의 약 22%를 차지한다.¹⁴⁾ 흡수·분포·대사·배설·독성(ADMET) 예측을 위하여 데이터를 학습에 사용해야 하지만 제약회사 간 데이터를 직접 공유하는 것은 보안 및 데이터 주권 문제 등 어려움이 존재한다.

연합학습(federated learning)은 여러 기관이 보유한 데이터를 물리적으로 공유하지 않고, 기계학습(ML) 모델 파라미터만 공유하여 데이터의 보호와 활용을 동시에 가능하게 하면서 인공지능(AI) 모델 성능을 전반적으로 개선하는 방법으로 국내·외에서 연합학습을 활용하여 비임상시험의 결과를 예측하기 위한 프로젝트가 완료·수행 중이다.

3.2.1. 유럽의 MELLODDY 프로젝트

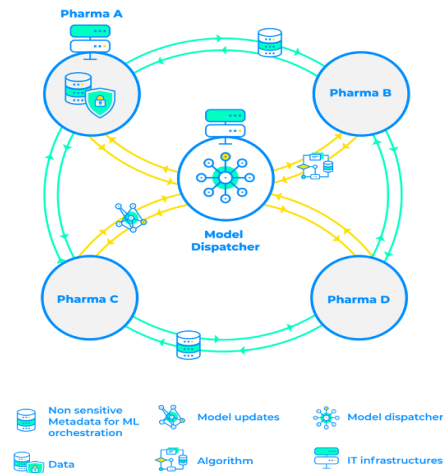
유럽에서는 연합학습을 사용하여 제약사들 간 협력 모델을 구축하고 신약 개발 가속화 및 효율성을 높이기 위한 ‘MELLODDY(Machine Learning Ledger Orchestration for Drug Discovery)’ 프로젝트를 운영하였다(‘19.6~‘22.5).¹⁵⁾ 본 프로젝트는 유럽연합(EU)과 제약업계의 공동 투자로 운영되었으며, 대형 제약기업 및 반도체 기업, 인공지능(AI) 신약개발 회사 등이 참여하였고, 기업이 보유한 정보는 보호하면서 다양한 기관의 많은 데이터를 연합학습하면 인공지능(AI) 모델의 예측 성능이 향상될 것이라는 가설 검증을 목적으로 진행되었다.

연합학습은 데이터셋을 교환하거나 데이터를 한 곳으로 모으지 않고 분산된 데이터셋에서 기계학습을 가능하게 하며, 분석 결과만을 중앙 서버로 전송하여 데이터의 보안 유지가 가능하다(그림 2). 본 프로젝트에서는 10개 제약회사의 약리 및 독성 실험 데이터를 학습시켰다. 이는 26억 개 이상의 데이터, 2,100만 개 이상의 화합물, 40,000개 이상의 실험을 문서화한 양이었으며, 개별 회사의 데이터 대비 약 10배에 달하는 수준이었다. 또한 공개적으로 사용 가능한 자료원의 데이터도 포함되었다.¹⁶⁾

14) Kraljevic S, et al. Accelerating drug discovery. EMBO Rep. 2004;5(9):837-42.

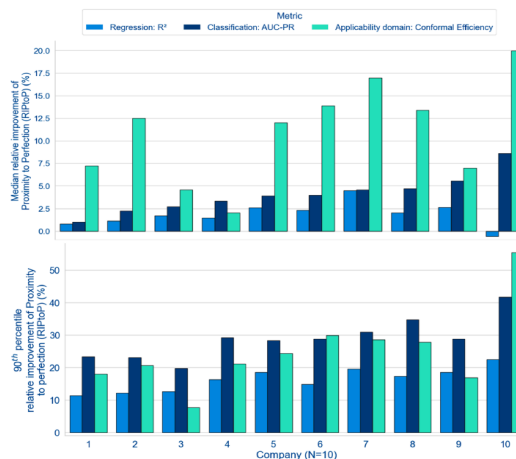
15) Heyndrickx W et al., MELLODDY: Cross-pharma Federated Learning at Unprecedented Scale Unlocks Benefits in QSAR without Compromising Proprietary Information. J Chem Inf Model. 2024;64(7):2331-2344.

16) Gaulton, A. et al., The ChEMBL Database in 2017. Nucleic Acids Res. 2017;45:D945-D954.

그림 2. EU-MELLODDY 플랫폼 개요¹⁷⁾

모델의 성능 평가를 위해 분류 모델(classification model)에서는 수신자 조작 특성 곡선 하 면적(area under the receiver operator characteristic curve, AUC-ROC) 및 정밀도 재현율 곡선 하 면적(area under the precision recall curve, AUC-PR)을 사용하였고, 회귀 모델(regression model)에서는 R^2 으로 평가하였다. 예측 모델의 성능 향상을 평가하는 지표로는 RIPToP(Relative Improvement of Proximity to Perfection)을 사용했으며, RIPToP 값이 100%에 가까울수록 모델이 완벽한 예측을 달성했음을 의미하며, 0%에 가까울수록 모델 성능에 차이가 없거나 개별 회사의 단일 모델과 동일한 성능을 가짐을 의미한다.

대부분의 회사에서 AUC-PR 및 R^2 의 RIPToP가 각각 4% 이상 및 2% 이상 향상되었으며, 이는 개발된 연합학습 모델이 더 넓은 범위에서 단일 모델보다 더 잘 작동할 가능성을 보여준다(그림 3). 본 프로젝트를 통해 경쟁 관계에 있는 제약회사들이 데이터를 공유하지 않고도 협력할 수 있는 새로운 모델을 제시하였고, 개별 회사의 독립적 모델보다 연합학습 모델에서 더 우수한 성능을 나타냈다.

그림 3. 각 제약회사 모델 대비 연합학습 모델의 성능 향상¹⁵⁾

17) Machine Learning Ledger Orchestration for Drug Discovery(<https://cordis.europa.eu/project/id/831472/reporting>) ('24.12.16. 기준)

3.2.2. 국내 K-MELLODDY 프로젝트¹⁸⁾

국내에서도 연합학습 기반 신약개발(Federated Drug Discovery, FDD) 플랫폼을 구축하고 흡수·분포·대사·배설·독성 예측 모델(Federated ADMET Model, FAM)을 개발하기 위한 ‘연합학습 기반 신약개발 가속화 프로젝트(K-MELLODDY)’를 추진·수행 중이다(‘24.4~‘28.12). 흡수·분포·대사·배설·독성 예측 모델(FAM)은 생체 외(*in-vitro*), 생체 내(*in-vivo*) 및 임상시험 데이터를 연계·사용하여 흡수·분포·대사·배설·독성(ADMET) 결과 및 임상시험 약동학 파라미터까지 예측하는 모델이다. 이를 통해 데이터 수집 및 공유 문제를 해결 할 수 있는 연합학습 운영 사례를 제시할 수 있을 것이라 예상된다.

3.2.3. 국내 제약사의 인공지능(AI)/기계학습(ML) 활용 사례

의약품심사소통단(CHORUS) 허가·심사 지원 분과의 ‘인공지능(AI) 활용 신약개발’ 소분과에 참여한 제약회사 조사 결과에 따르면, 일부 회사에서 내부 데이터를 사용하여 흡수·분포·대사·배설·독성(ADMET)을 예측하는 모델을 보유하고 있었다.

대분류	소분류	모델 유형
A	Caco-2 cell permeability	Regression
	HIA(Human Intestinal Absorption)	Classification
	Pgp inhibition	Classification
	Bioavailability(F)	Classification
	Lipophilicity	Regression
	Solubility	Regression
D	BBB permeability	Classification
	PPBR(Plasma Protein Binding Rate)	Regression
	VDss	Regression
M	CYP substrate	Classification
E	Half life	Regression
	Clearance (Microsome)	Regression
	Clearance (Hepatocyte)	Regression
T	LD50	Regression
	hERG inhibition	Classification
	Ames test	Classification
	DILI	Classification

18) K-MELLODDY 프로젝트(<https://kmelloddy.org/>, ‘24.12.16. 기준)

3.3. 임상시험

인공지능(AI)/기계학습(ML)은 임상시험에서 중재 연구(임상시험)와 비중재 연구(관찰연구)의 자료를 분석하여 대상자 모집 및 선정, 용법·용량 최적화 등에 활용될 수 있다.

최근 미국 국립보건원(National Institutes of Health, NIH) 산하 국립의학도서관(National Library of Medicine, NLM)과 국립암연구소(National Cancer Institute, NCI)의 연구팀은 대규모 언어모델(LLM) 기술을 활용해 환자에게 적절한 임상시험을 매칭할 수 있는 TrialGPT라는 프레임워크를 개발하였다¹⁹⁾. TrialGPT는 그림 4와 같이 3개의 모듈로 구성되어 있으며, TrialGPT-Retrieval 모듈에서는 대규모 분류를 통해 후보 임상시험을 검색하고, TrialGPT-Matching에서는 각 후보 임상시험에 대한 환자의 적합성 등을 평가한다. TrialGPT-Ranking에서는 Matching에서 평가한 정보를 통해 순위를 매겨 최종적으로 환자에게 적절한 임상시험에 대한 정보를 제공한다. TrialGPT를 활용해 환자가 특정 임상시험 기준을 만족하는지를 정확히 예측했는지 평가했을 때 임상 의와 유사한 수준의 정확도를 보였고, 환자 스크리닝에 소요되는 시간을 약 40% 단축시켰다.

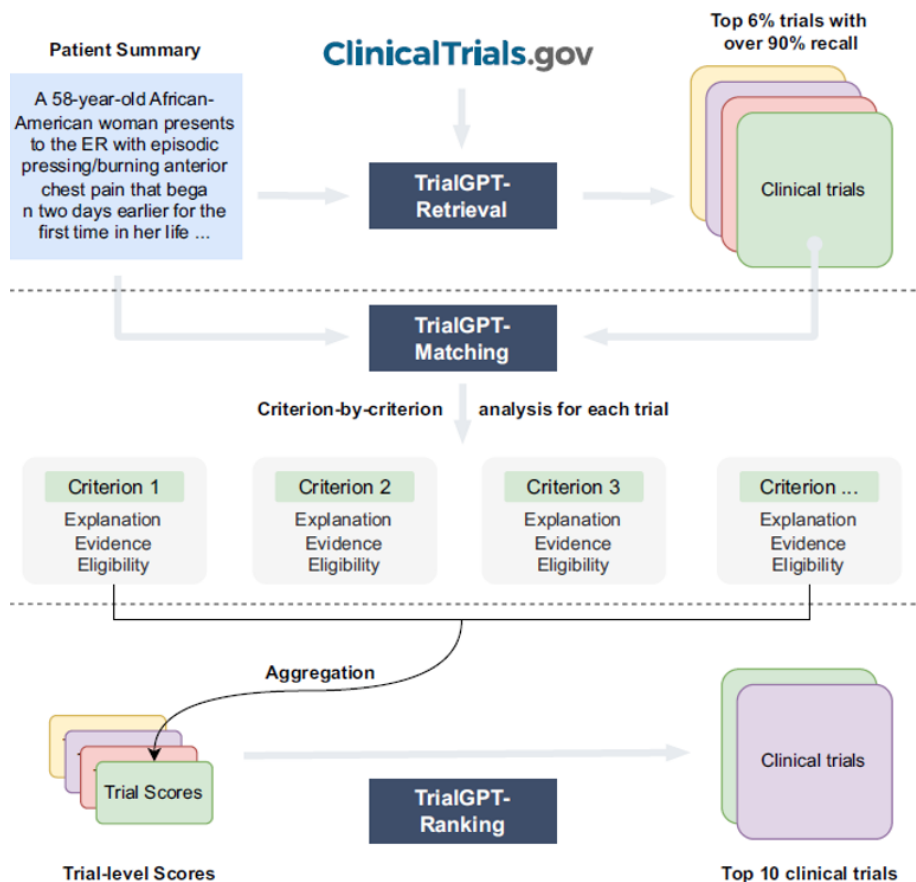


그림 4. TrialGPT의 구조¹⁹⁾

19) Jin Q et al., Matching patients to clinical trials with large language models. Nat Commun. 2024;15(1):9074.

인공지능(AI)/기계학습(ML)은 임상적 결과 예측을 위한 생체지표 발굴에도 활용될 수 있다. 미국 FDA는 노바티스와 호르몬 수용체 양성(HR+)/사람 상피세포 성장인자 수용체2 음성(HER2-) 전이성 유방암의 새로운 예후 및 예측 요인을 발견하기 위한 연구 협력을 하였으며, 본 프로젝트는 기계학습(ML)을 통해 임상 데이터, 유전체 데이터, 영상 데이터를 통합한 멀티모달(multimodal) 데이터를 분석하고, 이를 기반으로 환자의 임상 결과를 더 정확히 예측하여 정밀 의학을 지원하는 데 중점을 두며, 최근 통계 및 기계학습(ML) 분석 계획에 대해 문헌으로 발표되었다.²⁰⁾

최근 발표된 임상시험에 인공지능(AI)/기계학습(ML)이 활용된 사례를 분석한 연구 결과에 따르면, ClinicalTrials.gov(2023.03.01.기준) 데이터를 활용하여 ‘인공지능(artificial intelligence)’, ‘기계학습(machine learning)’, ‘딥러닝(deep learning)’ 키워드가 포함된 임상시험을 선별한 결과, 인공지능(AI) 관련 기술은 평가변수 평가(58.4%), 용법·용량 최적화(10.9%), 위험도 계층화(9.5%) 순으로 많이 활용되었으며, 평가변수 평가에서는 주로 이미지 및 비디오 분석을 통한 진단 또는 평가를 위해 활용되었다.²¹⁾

20) Multi-omics investigation on the prognostic and predictive factors in metastatic breast cancer using data from Phase III ribociclib clinical trials: A statistical and machine learning analysis plan (<https://www.medrxiv.org/content/10.1101/2023.08.30.23294367v1.full>)(‘24.12.16. 기준)

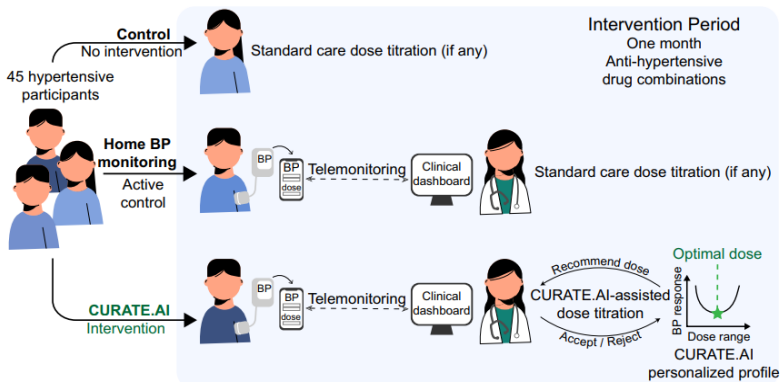
21) Go JM et al., Trends in Artificial Intelligence Applications in Clinical Trials: An analysis of ClinicalTrials.gov. Korean J Clin Pharm 2024;34(2):134-139.

아래 제시하고 있는 임상시험에 인공지능(AI)/기계학습(ML)이 활용된 사례는 의약품안전나라(의약품통합정보시스템)(<http://nedrug.mfds.go.kr>) 임상시험정보검색 및 ClinicalTrials.gov를 통해 확인하였다.

• 디지털 평가변수

임상시험 개요			
국가	<input checked="" type="checkbox"/> 국외 <input checked="" type="checkbox"/> 국내(다국가임상)		
품목명	- (타바파돈)	회사명	세레벨 테라퓨틱스
승인일	2023.11.30.	주성분명	타바파돈
임상시험 제목	파킨슨병에서 타바파돈의 58주 공개 임상시험(TEMPO-4 TRIAL)		
목적	파킨슨병 환자를 대상으로 타바파돈 장기 투여의 안전성 · 유효성 평가		
인공지능(AI)/기계학습(ML) 활용 내용			
항목	내용		
기타 평가변수	· 시험대상자의 개인 모바일 기기의 대화형 애플리케이션 플랫폼을 통해 시험대상자가 실시하는 발화 및 얼굴 표정을 원격 디지털 검사하여 수집한 오디오 및 비디오 자료의 탐색적 분석 (미국 시험기관 한정)		
참고자료			
· 의약품안전나라 임상시험정보('24.12.16. 기준) · https://clinicaltrials.gov/study/NCT04760769 ('24.12.16. 기준) · https://www.clinicaltrialsregister.eu/ctr-search/trial/2019-002952-17/IT ('24.12.16. 기준)			

• 용법 · 용량 최적화

임상시험 개요			
국가	<input checked="" type="checkbox"/> 국외 <input type="checkbox"/> 국내		
품목명	-	회사명	- (알렉산드라 병원(싱가포르))
NCT번호 (개시일)	NCT05376683 (2023.04.01.)	주성분명	-
임상시험 제목	고혈압 치료를 위한 AI 기반 약물 용량 조절		
목적	<p>CURATE.AI 기반 용량 조절 보조요법(원격 모니터링 플랫폼을 통한 일일 가정 혈압 모니터링과 CURATE.AI가 의사에게 용량을 추천하여 용량 조절을 지원하는 맞춤형 연속 용량 조절)의 활용 가능성 평가</p> 		
인공지능(AI)/기계학습(ML) 활용 내용			
항목	내용		
용법 · 용량 최적화	· 환자의 혈압 데이터를 사용하여 CURATE.AI 플랫폼을 활용한 개인별 항고혈압제 용량 조절		
참고자료			
· https://clinicaltrials.gov/study/NCT05376683 (‘24.12.16. 기준) · Truong ATL et al., CURATE.AI-assisted dose titration for anti-hypertensive personalized therapy: study protocol for a multi-arm, randomized, pilot feasibility trial using CURATE.AI (CURATE.AI ADAPT trial). Eur Heart J Digit Health. 2023;5(1):41-49.			

CURATE.AI 플랫폼²²⁾

- 개인 맞춤 약물 치료를 최적화하기 위한 인공지능 플랫폼으로, 환자별 데이터에 기반한 맞춤형 약물 용량 설정 및 임상 의사결정 과정을 지원하기 위한 목적으로 개발됨
- CURATE.AI를 활용하여 다양한 약물군(항암제, 항당뇨병제 등)에서 최적 용량 설정에 대한 연구가 수행됨 (NCT03759093, NCT03527238, NCT03832101, NCT02711956, NCT04522284, NCT05175235, NCT04769141 등)

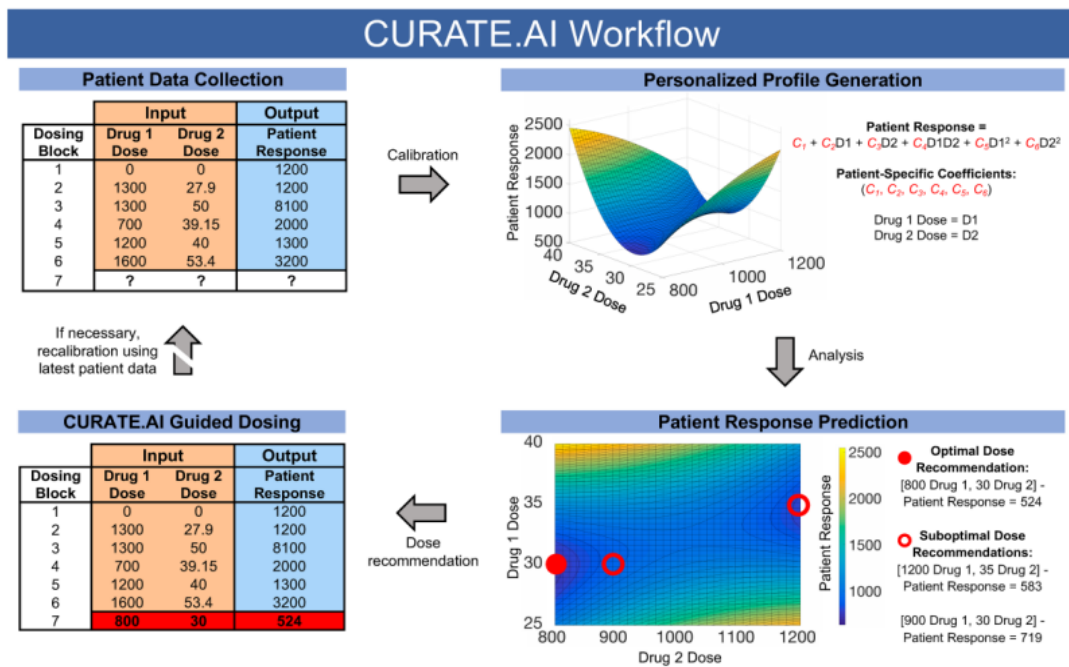


그림 5. CURATE.AI의 약물 용법·용량 최적화 과정

22) Blasiak A et al., CURATE.AI: Optimizing Personalized Medicine with Artificial Intelligence. SLAS Technol. 2020;25(2): 95-105.

• 위험도 계층화

임상시험 개요			
국가	<input checked="" type="checkbox"/> 국외 <input type="checkbox"/> 국내		
품목명	페마라	회사명	노바티스
NCT번호 (개시일)	NCT00241046 (2002.04.)	주성분명	레트로졸
임상시험 제목	호르몬 수용체 양성 유방암의 1차 및 2차 치료에서 레트로졸 사용 : 치료 전 위험 평가		
목적	전이성 유방암 환자에서 기계학습(ML)을 활용하여 특정 치료에 반응하는 환자와 반응하지 않는 환자를 90%의 정확도로 구별할 수 있는지 평가하고, 주어진 치료 하에서 질병 악화까지의 시간을 예측하며, 중기 생존 가능성이 높은 환자와 낮은 환자를 분류할 수 있는지 조사		
인공지능(AI)/기계학습(ML) 활용 내용			
항목		내용	
위험도 계층화		· 기계학습(ML)을 활용해 치료에 반응하는 환자와 반응하지 않는 환자를 예측하고 종양 성장 진행을 추정하여 생존기간 등 예측	
참고문헌			
· https://clinicaltrials.gov/study/NCT00241046 (‘24.12.16. 기준)			

• 대상자 선정

임상시험 개요			
국가	<input checked="" type="checkbox"/> 국외 <input type="checkbox"/> 국내		
품목명	칼퀸스 벤클렉스타	회사명	- (HOVON(벨기에, 네덜란드)와 Nordic(덴마크, 노르웨이, 스웨덴, 핀란드) 만성림프구성백혈병 연구 그룹 간의 협력 연구)
NCT번호 (개시일)	NCT03868722 (2019.10.11.)	주성분명	아칼라브루티닙, 베네토클락스
임상시험 제목	만성림프구성백혈병 국제학회(International Working Group on CLL, IWCLL) 치료 기준을 충족하지 않는 감염 고위험군 및/또는 초기 치료가 필요한 새롭게 진단된 만성림프구성백혈병(chronic lymphocytic leukemia, CLL) 환자에 대한 아칼라브루티닙과 베네토클락스 단기 병용치료		
목적	신규 진단된 만성 림프구성 백혈병 환자를 대상으로 한 단기 병용 치료(아칼라브루티닙 및 베네토클락스) 2상 임상시험 - CLL-TIM(기계학습 예측 알고리즘, 치료 감염 모델)을 통해 감염 고위험군 및/또는 초기 만성 림프구성 백혈병 치료가 필요한 환자를 식별한 후, 병용 치료(베네토클락스, 아칼라브루티닙) 3개월 동안 3등급 감염이 없고 치료가 불필요한(grade 3-Infection-free, treatment-free survival) 생존 기간이 개선될 수 있는지 평가함		
인공지능/기계학습(AI/ML) 활용 내용			
항목		내용	
대상자 선정		· 기계학습 알고리즘(CLL-TIM)을 통해 감염 고위험군 및/또는 초기 만성 림프구성 백혈병 치료가 필요한 환자 식별	
참고자료			
· https://clinicaltrials.gov/study/NCT03868722 (‘24.12.16. 기준)			

CLL-TIM(CLL Treatment-Infection Model) 모델²³⁾

- 4,149명의 CLL 환자 데이터를 기반으로 한 28개의 기계학습 알고리즘으로 구성된 앙상블 알고리즘으로, 실제 환경에서 예상되는 높은 비율의 결측 데이터를 포함한 이질적인 데이터(heterogeneous data)를 정밀도(precision) 72%, 재현율(recall) 75%로 처리하였음

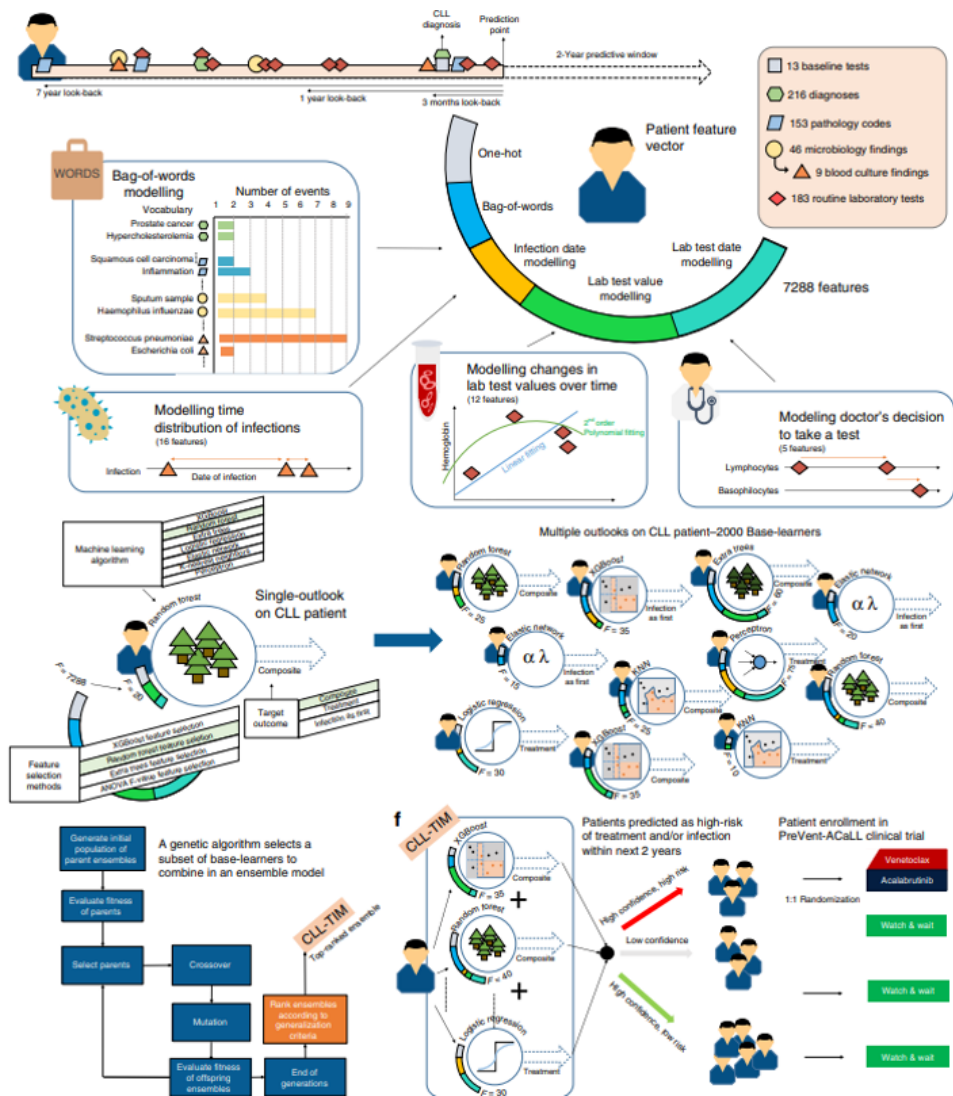


그림 6. CLL-TIM 개발 과정 모식도

23) Agius R et al., Machine learning can identify newly diagnosed patients with CLL at high risk of infection. Nat Commun. 2020;11(1):363.

임상시험 개요				
국가	<input checked="" type="checkbox"/> 국외 <input type="checkbox"/> 국내			
품목명	키너렛주 (KINERET® injection)	회사명	스웨디시 오픈 바이오비트럼	
NCT번호 (개시일)	NCT04357366 (2020.04.15.)	주성분명	아나킨라	
임상시험 제목	수용성 유로키나아제 플라스미노겐 활성화제 수용체(soluble urokinase plasminogen activator receptor, suPAR) 기반 아나킨라 치료를 통한 COVID-19로 인한 중증 호흡 부전의 위험 평가 및 초기 관리 검증: SAVE 공개, 비무작위, 단일군 임상시험			
목적	수용성 유로키나아제 플라스미노겐 활성화제 수용체 기반 중증 호흡 부전의 위험이 있는 COVID-19 폐렴 환자를 식별하고, 해당 환자에서 아나킨라를 투여하여 중증 호흡 부전의 진행을 막는 생체지표 도구로 활용			
인공지능/기계학습(AI/ML) 활용 내용				
항목	내용			
의약품 투여가 적절한 환자군 확인	<div>· 미국 FDA의 긴급사용승인 시 상용화된 수용성 유로키나아제 플라스미노겐 활성화제 수용체(suPAR) 측정 키트가 존재하지 않아, 미국 FDA에서는 인공지능(AI)/기계학습(ML)(엘라스틱넷 회귀분석(elastic net regression) 및 인공신경망(artificial neural network))을 활용하여 의약품 투여가 적절한 환자군을 확인하기 위한 점수 측정도구를 개발하였음</div> <div>· 다음 8가지 항목 중 최소 3가지 기준을 만족하는 환자는 기저치의 수용성 유로키나아제 플라스미노겐 활성화제 수용체(suPAR) 수치가 ≥ 6 ng/mL일 것으로 예측되었음</div>			
	<table><tr><th>측정도구</th></tr><tr><td>연령 ≥ 75세 WHO 기준에 따른 중증 폐렴환자 현재 또는 과거 흡연력 장기부전평가점수(sequential organ failure assessment score, SOFA score) ≥ 3 호중구-림프구 비율(neutrophil-to-lymphocyte ratio, NLR) ≥ 7 헤모글로빈 ≤ 10.5g/dL 허혈성 뇌졸중 병력 혈중 요소 ≥ 50mg/dL 및/또는 신장 질환 병력</td></tr></table>			측정도구
측정도구				
연령 ≥ 75 세 WHO 기준에 따른 중증 폐렴환자 현재 또는 과거 흡연력 장기부전평가점수(sequential organ failure assessment score, SOFA score) ≥ 3 호중구-림프구 비율(neutrophil-to-lymphocyte ratio, NLR) ≥ 7 헤모글로빈 ≤ 10.5 g/dL 허혈성 뇌졸중 병력 혈중 요소 ≥ 50 mg/dL 및/또는 신장 질환 병력				
참고자료				
<div>· https://clinicaltrials.gov/study/NCT03868722(‘24.12.16. 기준)</div> <div>· Liu Q, et al., Using Machine Learning to Determine a Suitable Patient Population for Anakinra for the Treatment of COVID-19 Under the Emergency Use Authorization. Clin Pharmacol Ther. 2024;115(4):890-895.</div>				

• 합성대조군(synthetic control arm) 활용

임상시험 개요			
국가	<input checked="" type="checkbox"/> 국외 <input type="checkbox"/> 국내		
품목명	- (MDNA55)	회사명	메디세나 테라퓨틱스
NCT번호 (개시일)	NCT02858895 (2017.04.11.)	주성분명	MDNA55
임상시험 제목	재발성 또는 진행성 교모세포종 성인 환자 대상 MDNA55의 전도 촉진 투여(Convection-Enhanced Delivery, CED)에 대한 공개, 비무작위, 다기관 2상 연구		
목적	재발성 또는 진행성 교모세포종 성인 환자 대상 MDNA55의 전도 촉진 투여 (Convection-Enhanced Delivery, CED)에 대한 2상 임상시험		
인공지능/기계학습(AI/ML) 활용 내용			
항목	내용		
합성대조군 활용	<div><div>· 단일군 2b상 임상시험 결과를 합성대조군(레지스트리 자료)과 비교하여, 탐색적으로 유효성을 평가하고 3상 임상시험에 합성대조군을 포함시키는 방법을 미국 FDA에 제안하였음. 미국 FDA는 무작위배정과 합성대조군을 모두 사용하는 하이브리드 방식을 수용하였음</div><div><div><div><div><div>Eligibility:</div><div><ul style="list-style-type: none">○ Age ≥ 18 yrs○ De novo rGBM○ IDH1/IDH2 WT○ KPS ≥ 70○ 1st or 2nd relapse○ Re-resection not indicated○ Tumor size <16 cm² (SPD)</div></div><div><div>Protocol enrollment into RCT</div><div>Protocol enrollment from ECA</div></div><div><div>Randomization 3:1</div><div>Propensity Score balancing</div></div><div><div>Bizaxofusp (N~150) Dose 240 µg</div><div>SOC (N~50)</div><div>SOC / Matched ECA (N~100)</div></div><div><div>Primary Endpoint<ul style="list-style-type: none">○ Overall survival (OS)</div><div>Secondary Endpoint<ul style="list-style-type: none">○ OS within IL4R subgroups○ OS within MGMT methylation subgroups○ OS-12○ Safety and tolerability</div></div></div></div></div></div>		
참고자료			
<div><div>· https://clinicaltrials.gov/study/NCT02858895(‘24.12.16. 기준)</div><div>· https://ir.medicenna.com/static-files/45900f52-929c-45ea-9d40-b860873de9ec</div></div>			

4

거버넌스 및 신뢰 기반 인공지능(AI)에 대한 업계 의견

인공지능(AI) 기술 발전과 더불어 윤리적 개발 및 활용을 위한 거버넌스 구축은 인공지능의 안전성과 신뢰성을 확보하는 데 있어 핵심 관심 대상이다. 인간 주도 인공지능(AI) 거버넌스는 신뢰할 수 있는 인공지능(AI)/기계학습(ML) 개발에 필수적인 책무와 투명성, 법적·윤리적 가치 등을 준수하는 데 도움이 될 수 있다. 특히 의약품 개발 시 활용되는 인공지능(AI) 기술의 안전하고 책임감 있는 사용은 중요하게 인식되어야 한다.

다음은 의약품 개발에서 인공지능(AI)/기계학습(ML) 활용 시 고려해야 하는 거버넌스 측면 등에 대해 업계가 바라보는 현재 수준의 관점 및 의견을 정리하였다. 의약품심사소통단 허가·심사 지원 분과의 ‘인공지능(AI) 활용 신약개발 소분과’에 참여한 제약회사를 중심으로 조사한 것으로, 제약업계 전체 의견을 대변하는 것은 아니다.

4.1. 거버넌스, 책임 및 투명성

Q1. 의약품 개발에서 인공지능(AI)/기계학습(ML) 활용 시 가장 규제 명확성이 필요한 분야는?

- ☞ · 임상개발 단계는 인공지능(AI)/기계학습(ML) 모델의 편향성과 불확실성으로 인하여 환자의 안전에 위험을 초래할 수 있으므로 데이터 완전성(integrity), 개인정보보호 등의 안전장치가 필요함. 특히 임상시험 설계와 관련하여 데이터 품질, 모델 투명성, 해석 가능성 및 규제 수용 기준 등에 대한 규제 명확성이 요구됨
- 인공지능(AI)으로 발굴한 생체지표의 인정 및 검증, 임상시험에서 유효성 평가 변수로 활용 등에 대한 증거 기반 입증 방법에 대한 규제 측면 안내가 필요한 영역임

Q2. 의약품 개발에서 인공지능(AI)/기계학습(ML) 활용 시 '투명성(transparency)'이 의미하는 바는?
(예: 투명성은 인공지능(AI)/기계학습(ML) 모델의 사용, 개발, 성능 및 가능한 경우에는 그 논리를 포함하여 모델에 대한 적절한 정보가 규제기관 또는 이해관계자에게 명확하게 전달되는 정도로 생각할 수 있음)

- ☞ 규제기관과 이해관계자가 인공지능(AI)/기계학습(ML) 모델의 개발, 사용, 성능 및 논리를 이해하기 위한 인공지능(AI)/기계학습(ML) 모델 관련 모든 정보(한계점 및 잠재적 위험 포함)에 대한 명확한 의사소통을 의미한다고 볼 수 있음

Q3. 인공지능(AI)/기계학습(ML) 활용 의약품 개발 과정에서 경험한 투명성을 방해하는 주요 장벽 요인은 무엇인가?

- ☞ 데이터 진위 여부, 데이터 제한(학습데이터 한계에 따른 과적합 발생, 긍정적인 데이터만 이용하는 경우 등), 모델 가중치(다양한 해석 가능), 모델 복잡성(인공지능(AI)/기계학습(ML), 특히 딥러닝 모델은 복잡하고 불투명하여 의사결정 과정을 완전히 설명하기 어려움)

Q4. 추적가능성(traceability) 및 점검가능성(auditability)을 개선/향상시키기 위해 계획 중(또는 수행 중)인 절차가 있는가?

- ☞ 데이터 및 모델의 포괄적인 문서화, 체계적인 버전 관리, 표준지침 마련, 로그 관리(자동 로깅(logging) 시스템 구축 등)

Q5. 의약품 개발에서 인공지능(AI)/기계학습(ML)의 안전하고 효과적인 사용을 위해 사전에 규정된 관리 방법과 변경 사항을 모니터링하고 수집하는 방법이 있는가?

- ☞ 인공지능(AI)/기계학습(ML) 모델의 목표/범위/요구사항 등을 명확히 정의하고, 데이터 전처리 과정을 명확하게 정의하고 문서화해야 함. 적절한 알고리즘을 선택하고, 알고리즘 선택 사유 및 제약사항 등을 문서화함. 이때 코드와 아티팩트(artifact)에 대한 버전 관리를 시행하여 변경 사항을 추적할 수 있도록 함. 또한 변경 관리 시스템 등을 사용하여 변경 사항을 기록하고 변경 사항에 대한 영향을 평가하여 모델 성능에 미치는 영향을 평가함

4.2. 데이터 관련 고려사항(품질, 신뢰성, 대표성)

Q1. 인공지능(AI)/기계학습(ML) 활용 의약품 개발 시 추가적인 데이터에 대한 고려사항에는 어떤 것이 있는가?

- ☞ · 데이터 품질 및 완전성(integrity) : 사용되는 데이터가 일관되고 정확한지 여부, 모델 훈련과 검증에 필요한 모든 관련 변수를 포함하는지 여부, 결측치/이상치/노이즈 전처리 여부 등 (예: 적절한 결측치 처리 방법 사용 필요)
- 데이터 접근성 및 다른 데이터베이스와의 ID 체계 호환성
- 데이터 보안 및 개인정보보호, 데이터 관리 및 추적

Q2. 인공지능(AI)/기계학습(ML)의 완전성(integrity) 보장 방법 또는 비둘림, 결측치, 그 외 데이터 품질과 관련된 고려사항과 같은 문제점을 해결하기 위해 사용하고 있는 방법/절차에는 어떤 것이 있는가?

- ☞ · 결측치 처리, 중복 제거 및 데이터 형식 표준화를 위한 데이터 정리, 전처리단계 필요. 정확하고 일관된 데이터 라벨링(예: 라벨링 프로세스 자동화), 자료원의 신뢰성과 정확성을 보장하기 위한 철저한 검증 필요
- 모델의 성능과 일반화 능력을 평가하기 위해 교차 검증 기법(예: k-fold 교차 검증)을 적용하고, 외부 데이터에 대한 예측 성능 평가를 위해 밸리데이션(외부타당성) 수행

Q3. 개인정보보호 및 보안을 향상시키기 위해 주요 방법/절차에는 어떤 것이 있는가?

- ☞ 데이터 암호화, 데이터 접근 권한 설정(전산 보안, 접근 제한, 로그기록 작성 등)

Q4. 재현성(reproducibility) 및 반복성(replicability) 문제를 해결하기 위해 사용하는 주요 방법/절차에는 어떤 것이 있는가?

- ☞ · 데이터 수집, 전처리, 모델 개발 및 검증의 모든 단계에 대한 상세한 SOP 마련
- 데이터 수집 방법 차이에 따른 변동성을 최소화하기 위해 표준화된 데이터 수집 방법과 기준 사용
- 자료원 및 전처리, 모델 아키텍처, 하이퍼파라미터, 훈련 절차 등 모든 프로세스 문서화
- 버전 관리 시스템(예: Git, GitHub 등)을 사용하여 코드 및 모델의 변경사항 추적
- 표준화된 벤치마크 및 공개 데이터셋을 사용하여 모델 성능 비교

Q5. 비둘림 식별 및 관리를 위해 어떤 방법/절차를 사용하는가?

- ☞ · 데이터 특성 및 분포를 확인하고 잠재적 편향을 식별하기 위해 탐색적 데이터 분석(EDA) 수행
- 모델 훈련에 사용되는 데이터를 다양한 자료원으로부터 충분히 수집하여 데이터 대표성 확보
- 편향을 줄이고 모델이 최신 상태로 유지되도록 정기적으로 업데이트된 데이터를 사용하여 모델 재훈련, 불균형 데이터 처리(오버/언더 샘플링), 피쳐 엔지니어링

4.3. 모델 관련 고려사항(모델 개발, 성능, 모니터링 및 밸리데이션)

Q1. 의약품 개발 시 적용할 수 있는 인공지능(AI)/기계학습(ML) 모델 개발 및 성능에 대해 문서화하고 있는가?

- ☞ 모델 개발 시 사용한 데이터, 아키텍처, 성능을 기록하여 관리, 버전 관리 시스템(예: Git, GitHub 등), 모델 학습에 사용한 스크립트를 python(.py), Jupyter Notebook 등의 문서로 관리하기도 함

Q2. 사용 용도에 적절하게 모델 유형과 알고리즘을 선정하였는가?

- ☞ 목적과 의미에 맞는 알고리즘을 선정하여 사용

(예: AutoML Tool을 사용하여 적절한 알고리즘 선정에 활용, EDA Tool을 사용하여 데이터 특성에 맞는 알고리즘 선정에 활용, 필요한 경우 앙상블 알고리즘으로 모델 성능 향상에 활용, 단백질 서열 정보를 반영하기 위한 대규모 언어모델(LLM), 화합물의 구조를 반영하기 위한 그래프 모델 적용 등)

Q3. 모델 활용 단계에서 모델을 검증하고 성능을 측정하기 위해 특정 접근법을 언제 사용할지 결정하는가?

- ☞ 모델의 종류(분류모델/회귀모델)에 맞는 평가지표(metric)를 사용하고, 분류모델의 경우 데이터의 불균형 여부 등에 따라 적절한 평가지표를 선정함. 표준 벤치마크 데이터가 있는 경우, 해당 데이터를 사용하여 모델의 성능을 다른 모델과 비교·평가함

Q4. 투명성과 설명 가능성을 평가하고 모델 투명성을 높이는가?

- ☞ · 데이터 수집, 전처리, 알고리즘 선택, 하이퍼파라미터 튜닝, 성능 평가 등 모든 단계를 문서화함
- 인공지능(AI)/기계학습(ML) 모델의 설명 가능성 및 투명성을 높이기 위해 SHAP 등 각 특성이 모델 예측에 기여하는 정도를 정량화할 수 있는 도구로 사용할 수 있음

Q5. 정확성 및 설명가능성 문제를 해결하는가?

(예: 모델의 정확도는 향상시킬 수 있으나, 설명가능성에 제한점이 있는 경우 등)

- ☞ 모델의 성능 향상 및 설명가능성을 부여할 수 있는 모델로 설계. 다만, 항상 설명할 수 있는 결과가 도출되지는 않음

Q6. 인공지능(AI)/기계학습(ML) 모델 개발을 위해 공개된(open-source) AI 소프트웨어 선택 여부 및 공개 소프트웨어 사용 시 고려사항은 무엇인가?

- ☞ · 공개된 소프트웨어를 사용하는 경우 적절한 알고리즘인지, 확장 가능성이 있는지, 사용된 패키지 버전 등 확인
- 필요에 따라 공개된 소프트웨어를 사용하기도 하며, 성능, 정확성 및 안정성을 평가함(활발히 유지·관리되는지, 버그와 취약점에 대한 패치가 제공되는지 확인)
- 모델 개발을 위해 Tensorflow, PyTorch, Scikit-learn 등 공개된 프레임워크 사용. 공개된 프레임워크 사용 시 해당 프레임워크의 커뮤니티 활동과 문서화 수준을 확인하고, 해당사항이 있는 경우 라이선스를 검토하여 제약사항이 있는지 확인함

Q7. 자료원 선정과 선정/제외기준을 기록하고 이에 대한 정보를 제공하기 위해 어떤 방법/절차를 사용하고 문서화하는가?

- ☞ 데이터 소스를 선택하기 전에 인공지능(AI)/기계학습(ML) 모델 활용 목적을 명확히 해야 함. 활용 목적에 맞도록 여러 데이터 소스를 평가하여 적합성을 평가하고, 데이터 선택 과정, 평가 기준, 선정/제외 기준, 최종 선택된 데이터 소스에 대한 상세 정보를 문서화하여야 함

Q8. 어떤 사용 맥락에서 설명가능성을 다루고 있으며, 성능과 설명가능성에 대한 고려사항의 균형을 어떻게 조절하는가?

- ☞ · 복잡한 모델 대신 상대적으로 간단한 모델을 사용하여 설명가능성을 높일 수 있음(예: 신경망 모델과 결정트리 모델의 성능이 비슷한 경우, 복잡한 신경망 대신 설명 가능한 결정트리 모델을 사용할 수 있음). 성능과 설명가능성 간의 가치교환을 평가하여 최적의 균형을 찾도록 함
- 모델에서 예측된 결과를 해석하고 설명하는데 사용(예: 화합물의 어떤 구조로 인해 독성이 발생했는지, 모델의 성능이 낮을 때는 왜 낮은지, 성능이 높을 때는 왜 높은지 설명가능성으로 판단하여 기존 알고리즘 대비 성능 개선 정도 확인)

Q9. 모델 예측에서 불확실성 평가를 문서화하기 위해 어떤 접근법을 사용하고 있으며 불확실성을 어떻게 전달하는가? 불확실성 평가를 지원하기 위해 어떤 방법과 표준을 개발해야 하는가?

- ☞ 시각화 도구(신뢰구간 플롯, 히스토그램, scatter plot 등)를 사용하여 불확실성 정보에 대한 이해를 도울 수 있으며, 분류모델 예측 결과 전달 시 예측값과 함께 예측 확률값을 제공하여 예측의 신뢰도를 나타낼 수 있음

인공지능(AI) 활용 의약품 개발 사례집

발행일	2024년 12월 24일
발행인	강석연
편집위원장	김영림
집필위원	식품의약품안전평가원 의약품심사부 순환신경계약품과 김소희, 주정훈, 서현옥, 변지영, 김지명, 김정현, 현양진, 배수영, 조혜영, 김연재(청년인턴)
도움주신분	의약품 심사 소통단(CHORUS) '인공지능(AI) 활용 신약개발 소분과' (AI신약융합연구원, 한국의약품수출입협회, 대웅제약, 삼진제약, JW중외제약, 한미약품)
문의처	식품의약품안전평가원 의약품심사부 순환신경계약품과

인공지능(AI) 활용 의약품 개발 사례집



공직자 부패·공익 신고안내 *신고자 및 신고내용은 보호됩니다.
식약처 대표누리집 “국민소통 → 여론광장 → 부패·공익신고” 클릭