NOVARTIS



발견

노바티스의 디지털 연구개발 플랫폼 data42, 치료의 새로운 정의를 세우다

Oct 30, 2020

노바티스는 data42를 바탕으로 한 데이터 과학을 활용하여, 의약품 연구개발의 새로운 정의를 세워가고 있습니다. *#AI #reimaginingmedicine*

인공지능은 수십 년 동안 존재해왔지만, 최근 '딥러닝(deep learning)'의 발전으로 데이터 과학자들은 놀라운 도약을 경험할 수 있게 되었습니다. 데이터 입력에 의해 학습하는 기존의 머신 러닝 알고리즘과 달리, 딥러닝은 알고리즘을 촘촘히 적용해 스스로 학습하고 지적 의사결정을 할 수 있는 '인공 신경 회로망'을 만들었습니다.

딥러닝 분야의 선두주자 중 하나인 '딥마인드(DeepMind)'는 약 3년 전 매우 복잡한 룰이 적용되는 '바둑'에서 인간을 이겼고, 유사한 인공지능 접근법은 2019년 단백질 형태를 예측하는데 있어서 생물학자들의 두뇌를 추월했습니다.

이러한 발전은 연구개발에 새로운 길을 열어주었습니다. 이는 현 시대 의학 및 운영상의 가장 큰 난제들을 해결하는데 도움이 될 수 있습니다. 계속 상승하는 연구개발비를 관리하는 것은 물론, 적절한 치료법이 없거나 제한적인 질환에 대한 혁신적인 치료법 개발 등이 그것입니다

도전은 정말 중요합니다. 새로운 치료제가 시장에 출시되려면 20억 달러 이상의 투자가 필요하며, 평균 10년 이상의 연구기간이 소요됩니다. 통상적으로 임상연구 단계의 분자물질 10개 당 1개만이 시장에 나옵니다.

“모든 데이터가 수집되면, 새로운 통찰력을 창출할 수 있는 잠재력은 엄청날 것입니다.”

의학적 측면에서도 '딥러닝'에 대한 수요는 증가하고 있습니다. 지난 10년 동안 미국에서 약 500개의 신약이 승인되었음에도 적절한 치료법이 없는 분야의 의학적 수요는 여전히 높습니다. 예를 들어, 알츠하이머와 같은 만성 · 퇴행성 질환은 여전히 치료가 어려우며, 7,000개가 넘는 희귀질환 대부분은 혁신적인 치료 옵션이 없는 상황입니다.

data42의 테크놀러지 리더(Technology Leader)인 파스칼 부케(Pascal Bouquet)는 이렇게 말합니다. "흔히 새로운 데이터 및 디지털 기술을 사용할 때 우리가 무엇을 찾아낼 수 있을지 예측할 수는 없습니다. 그러나 곧 새로운 통찰력을 발견할 수 있을 것입니다. 지금까지 경험하지 못한 가치를 발견하고, 장기적으로는 순전히 데이터 기반의 신약을 완벽하게 설계하고 개발할 수 있을 것이라고 확신합니다”

이러한 가능성으로 제약사들은 모두 디지털 전문기술을 강화하고 있으며, 전통적인 의약품 개발모델을 뛰어넘는 혁신적인 치료법 개발에 대한 희망을 갖고 구글, IBM, 애플과 같은 기업들을 헬스케어 분야로 끌어들이고 있습니다.

미국글로벌시장조사업체 피치북(PitchBook)에 따르면, 2018년 벤처투자자들이 헬스케어 관련 인공지능 스타트업에 투자한 금액이 약 10억 달러를 상회하는 것으로 보고되었습니다. 그리고 그 시장은 점점 더 뜨거운 감자가 될 것입니다. 컨설팅&리서치 전문기업인 에베레스트 그룹(Everest Group)은 인공지능 기술에 대한 전반적인 헬스케어 투자가 2017년 15억 달러에서 2020년에는 60억 달러 이상으로 증가할 것이라고 예상합니다.

200만 환자년(patient-year)* 데이터 자산

노바티스는 이러한 신생 분야를 선도하고 있다고 믿습니다. 파스칼 부케는 "우리는 약 200만 환자년의 데이터를 보유하고 있습니다. 이것은 우리가 인공지능을 활용해 데이터를 선별하고, 지금까지 알려지지 않은 치료제와 질병 간의 상관관계를 찾는데 있어서 앞으로 나아가야 할 중요한 자산입니다."라고 설명합니다.
*환자년(patient-year): 각 환자 당 서로 다른 관찰기간의 합

이러한 비전을 실현하기 위해서는 먼저 모든 임상 및 연구 데이터(잠재적인 리얼월드 데이터, 영상 및 센서 데이터)를 구성하고 단일 플랫폼으로 취합하는 이른바 '데이터 레이크(data lake)*'를 만들어야 합니다. 그러나 개별 데이터세트는 종종 성별, 연령, 가족 및 질병 정도와 같은 측정점을 표기하기 위해 서로 다른 매개 변수를 사용하기 때문에 실제 실행하는 것은 말처럼 쉽지 않습니다.
*데이터 레이크(data lake): 가공되지 않은 상태로 저장되어 접근 가능한, 방대한 양의 데이터

data42의 제품 개발 리더(Product Development Lead) 피터 스피어(Peter Speyer)는 "수집된 모든 데이터는 머신 러닝이 가능하도록 정리되고 다듬어져야 합니다. 이는 수고롭고 번거로운 작업이지만, 그렇게 해야 데이터 과학자들이 데이터를 활용해 문제의 답을 찾는데 집중할 수 있습니다." 라고 말합니다.

데이터의 용량은 어마어마합니다. 연구 개발에 입력되는 데이터만 해도 20페타바이트(petabyte)로, 이는 MP3 플레이어에서 약 40,000년 동안 재생할 수 있을 정도의 음악파일 용량에 해당합니다.

"우리의 시스템에는 약 200만 환자년의 데이터가 있습니다. 이것은 우리가 인공지능을 활용해 데이터를 선별하고, 지금까지 알려지지 않은 치료제와 질병 간의 상관관계를 찾는데 있어서 앞으로 나아가야 할 중요한 자산입니다."

가치 있는 데이터 발굴

노바티스 생명의학연구소(NIBR, Novartis Institute for BioMedical Research)와 그룹 신약연구개발(GDD, Global Drug Development) 사업부, 비즈니스 서비스(NBS, Novartis Business Services) 사업부를 통틀어 100명 이상이 참여하는 data42 프로그램은 지금까지 큰 성과를 이루었습니다. 2,000건 이상의 임상 연구를 플랫폼에 도입했고, 데이터 깊숙이 저장된 새로운 정보를 찾는데 도움이 되는 머신러닝 모델을 12개 이상 테스트했습니다.

data42 프로그램 운영진은 사업 모델의 실효성과 객관성 지표를 확보하기 위해 비즈니스 결과를 중심으로 매우 구체적이고 명확한 단기 목표를 설정했습니다. 이를 반영하여 진행하고 있는 최근 프로젝트 중 하나는 류마티스 관절염의 생물학적 특성을 기반으로 질환의 세부 유형을 판별하는 프로젝트입니다.

피터 스피어는 "이 프로젝트를 위해 우리는 해당 질병 영역에서 우리의 기존 실험 데이터를 정리하는 작업을 진행하고 있습니다.”라고 합니다. 또한 그는 “우리의 목표는 치료법 중 하나에서 높은 반응을 보이는 하위 그룹을 식별하는 것입니다. 만약 이것이 가능하다면, 해당 사업부는 잠재적으로 새로운 임상연구를 고안하고 그 발견한 사실들을 테스트할 수 있을 것입니다."라고 설명합니다.

질문을 바탕으로 한 사고

현재 data42 프로그램 팀은 특정 암에서의 질병 진행에 대해서도 연구하고 있습니다.

또한, 이 팀이 데이터를 세밀하게 조정하고, 지금까지 사람들의 관심 밖이었던 질병 관련 정보를 찾아낼 거대한 데이터 레이크를 구축하는 중이기 때문에, 지금도 가치 있는 정보들은 보다 많이 발굴되고 있습니다.

피터 스피어는 “모든 데이터가 수집 및 정리되고 나면, 새로운 통찰력을 볼 수 있는 가능성은 엄청날 것입니다. 심부전을 예로 든다면, 심부전이 관심 질환으로 식별되든 동반질환이나 이상반응으로 파악되든 어떤 질문에 대해서도 해당 데이터를 놓고 그 관계를 분석해 볼 수 있을 것입니다. 이것이 바로 data42의 확장성입니다.”라고 말합니다.

data42프로그램이 이러한 가능성에 충족된다면, 데이터 과학자가 연구실이나 진료실에서 일하는 과학자들과 협업하는 방식도 변할 것입니다. "데이터를 준비하는 데이터 과학자와, 질문을 이해하고 데이터에서 무엇을 뽑아내야 하는지 고민하는 의학자들 간의 협업이 보다 강화될 수 있습니다." 라고 파스칼 부케는 설명합니다.

생물학자들과 화학자, 의사 모두 새로운 디지털 툴에 의해 대체되겠지만, 이 데이터 분석 툴은 데이터가 입력되면 될 수록 그 역할을 훌륭하게 해낼 것입니다. data42의 수장 아킴 플뤼케바움(Achim Pluekebaum)은 "질문을 잘 만들면, 생각만큼 해결책이 복잡하지 않다는 것을 알게 됩니다."라고 말합니다. "모든 질문에 이 인공지능이 모두 필요하다는 것은 아닙니다. 어떤 질문들은 통계학을 적용해 올바른 데이터를 찾을 수 있습니다. 올바른 방법을 적용하면, 답이 나옵니다. 질문을 통해 생각한다는 것은 인공지능의 유무와 관계없이 통찰력을 가속화하고 개선하는데 유용한 접근방법이 될 것입니다."

# # #

노바티스는 디지털 연구 개발 플랫폼인 data42 자산을 활용하기 위해 인간의 호기심과 컴퓨터 전문기술에 총력을 기울이고 있습니다.

Source URL: *https://www.novartis.com/kr-ko/stories/discovery/newsletter-9*

List of links present in page

- https://www.novartis.com/kr-ko/kr-ko/stories/discovery/newsletter-9
- https://www.novartis.com/kr-ko/kr-ko/stories/discovery