한국빅데이터학회지
제8권 제2호, 2023, pp. 125-131 https://doi.org/10.36498/kbigdt.2023.8.2.125

프라이버시와 개인화를 위한 고품질 표준 데이터 기반 약물감시 시스템 연구

High-Quality Standard Data-Based Pharmacovigilance System for Privacy and Personalization

양세모 · 송인서 · 이강윤†

가천대학교 IT융합공학과 컴퓨터공학부

요 약

전 세계적으로 약물 부작용은 주요 사망원인의 상위를 차지하고 있다. 약물 부작용에 대해 효과적으로 대응하기 위해, 능동적인 실시간 분석 기반 약물감시 체계로의 전환과 함께 데이터의 표준화와 품질 향상이 필요하다. 이를 위해, 개별 기관의 데이터를 통합하고 대규모 데이터를 활용하여 약물 부작용 예측의 정확도를 높이는 것이 중요하다. 하지만, 각 기관 간의 데이터 공유는 프라이버시 문제를 야기시키고 각기 다른 데이터 표준 구성도 다르다. 본 연구에서는 이 문제를 해결하기 위해, 개인정보보호 법규에 따라 데이터를 직접 공유하지 않고 모델의 학습 결과를 공유하는 연합학습 방식을 채택한다. 각 기관마다 다른 데이터 포맷을 Common Data Model(CDM)을 활용하여 데이터 표준화를 수행하고 데이터의 정확성과 일관성을 확립한다. 또한, 클라우드 기반의 연합학습 환경을 구성하여 보안 및 확장 관리에 효율성을 높이는 약물감시 시스템을 제안한다. 이를 통해 기관 간 데이터의 프라이버시를 보호하면서도, 효과적인 의약품 부작용 모니터링과 예측이 가능하다. 약물 부작용으로 인한 사망률 감소와 의료비용 절감을 목표로 하며, 이를 실현하기 위한 다양한 기술적 접근과 방법론을 탐구한다.

■ 중심어 : 약물감시, 연합학습, 공통데이터모델, 프라이버시, 클라우드

Abstract

Globally, drug side effects rank among the top causes of death. To effectively respond to these adverse drug reactions, a shift towards an active real-time monitoring system, along with the standardization and quality improvement of data, is necessary. Integrating individual institutional data and utilizing large-scale data to enhance the accuracy of drug side effect predictions is critical. However, data sharing between institutions poses privacy concerns and involves varying data standards. To address this issue, our research adopts a federated learning approach, where data is not shared directly in compliance with privacy regulations, but rather the results of the model's learning are shared. We employ the Common Data Model (CDM) to standardize different data formats, ensuring accuracy and consistency of data. Additionally, we propose a drug monitoring system that enhances security and scalability management through a cloud-based federated learning environment. This system allows for effective monitoring and prediction of drug side effects while protecting the privacy of data shared between hospitals. The goal is to reduce mortality due to drug side effects and cut medical costs, exploring various technical approaches and methodologies to achieve this.

■ Keyword : Pharmacovigilance, Federated Learning, Common Data Model, Privacy, Cloud

2023년 11월 24일 접수; 2023년 12월 11일 수정본 접수; 2023년 12월 18일 게재 확정.

* 본 연구는 과학기술정보통신부의 재원으로 과학기술일자리진흥원의 연구산업육성사업(grant number: 2022-미래연구서비스개발 지원-1-SB4-1)과 한국연구재단의 기초연구사업 (grant number: NRF-2022R1F1A1069069) 지원에 의하여 수행하였습니다.

† 교신저자 (keylee@gachon.ac.kr)

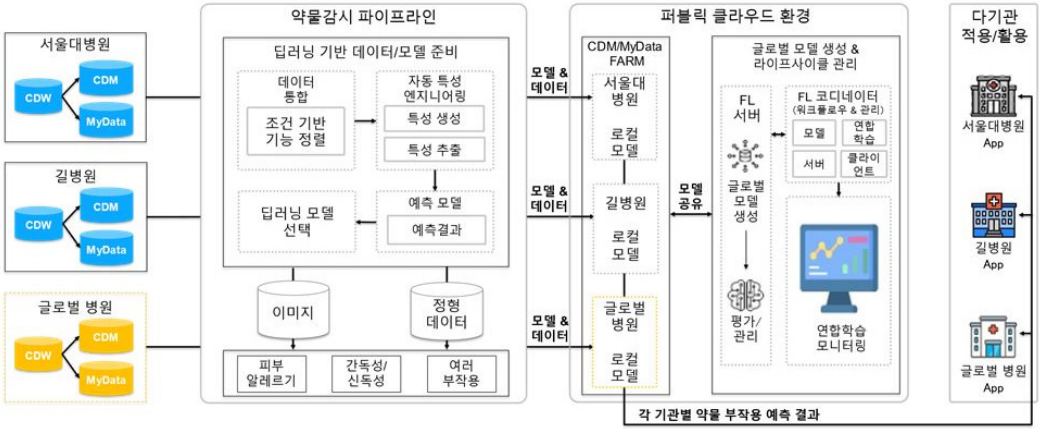
Ⅰ. 서론

약물 부작용은 전 세계에서 주요 사망원인의 상위를 차지하고 있으며, 미국에서는 연간 200만 례의 약물유해반응이 발생하고 10만 명이 사망하고 있다. 국내 3차 병원에서 약물 유해반응 발생률은 100명의 퇴원 환자 당 9.6건이며 총 입원 환자 중 1.7%가 약물 유해반응으로 인해 입원한다.[1] 시판 후 약물 유해정보 검색에서 중요한 역할을 하는 국내의 자발적 부작용 신고제도는 인식상의 문제, 과소 보고의 문제, 신뢰할 수 없는 정보, 자발적 보고의 낮은 질 등의 한계가 지속적으로 제기되며 약물 부작용 감시에 더 효과적인 시스템이 필요하다.

현재 약물 부작용 감시는 수동적인 자발적 보고자료 의존 체계에서 능동적인 실시간 분석 체계로 발전하고 있다.[2] 능동적인 감시를 통해 Electronic Medical Record(EMR)[3], CDM[4] 같이 다른 목적으로 축적된 데이터를 활용하여 시간과 비용을 절약할 수 있으며, 사용이 드문 의약품과 잘 발생하지 않는 유해사례도 확인가능하고 정확한 발생률 추정이 가능하다. CDM 기반의 약물 부작용 모니터링과 이상치 검출은 각종 이상사례를 신속하고 체계적으로 수집, 평가하여 의약품과 관련된 피해 발생을 예방하기 위한 활동으로 시판 후 안전관리에 있어 핵심적이다.

약물 감시 과정은 관련 데이터를 수집하는 것부터 시작된다. 하지만, 관련 데이터를 수집하려면 각 병원들 간의 데이터를 공유해야 하는데 개인정보 보호로 인하여 병원 간의 데이터를 쉽게 공유할 수 없다는 점이 현실의 한계이다. 본 연구에서는 이러한 한계를 극복하기 위해, 각 기관은 데이터를 직접 공유하지 않고 개별적으로 학습한 인공지능 모델의 결과를 공유하는 연합학습 기술을 활용하였다. 각 기관들은 데이터 공유 없이 프라이버시를 지키며, 표준화된 약물감시에 대한 공통 모델과 연합학습을 활용하여 환자의 개인정보를 보호한 채로 효과적인 의학적 결정을 내릴 수 있다.[5,6] 이러한 협력적이고 안전한 방식은 방대한 데이터를 통해 신속한 약물감시가 가능하게 함과 동시에 병원 간의 환자 개인정보 보호를 동시에 보장할 수 있다.

본 연구에서 제안하는 시스템은 퍼블릭 클라우드 기반의 연합학습 환경에서 운영하여 약물감시 시스템의 보안 및 확장 관리에 이점을 가진다. 약물 감시 시스템이 다양한 규모의 기관들과 원활하게 협력할 수 있게 하며, 새로운 기관이 시스템에 참여하거나 기존 기관의 데이터가 증가함에 따라 시스템을 쉽게 확장할 수 있다. 또한, 클라우드 기반 시스템은 각 기관이 필요에 따라



〈그림 1〉 클라우드 환경에서의 약물감시 연합학습 시스템 아키텍처

컴퓨팅 자원을 조정할 수 있어, 효율적인 자원 관리가 가능하며 비용 절감이 가능하다. [7] 퍼블릭 클라우드 기반의 연합학습 환경은 약물 감시 시스템의 성능 향상뿐만 아니라, 광범위한 네트워크를 통한 다양한 데이터의 신속한 분석과 처리가 가능하다. 이는 약물 부작용 예측의 정확성을 높이고, 실시간 모니터링을 통해 약물 관련 위험을 빠르게 식별하여 대응할 수 있다.

Ⅱ. 관련 연구

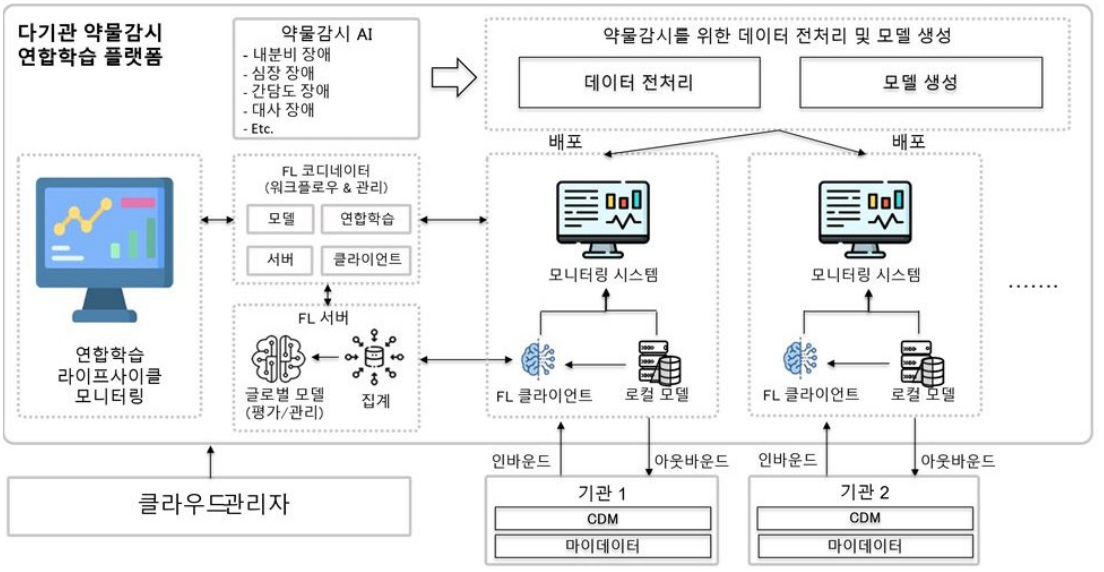
2.1 표준 CDM 기반 약물감시 시스템

개인 맞춤형 약물사용 스마트 도우미 서비스인 PharmCoach는 환자용 의사결정도우미 시스템이다. PharmSafe와 헬스아바타 플랫폼을 활용하여 스마트폰 및 웨어러블 기기로부터의 운동량, 식습관, 감정, 혈압 등의 라이프로그 데이터와 연계하고 환자용 약물사용 도우미 스마트 앱 서비스를 제공한다.[8] 국민건강보험심사평가원에서는 의약품 복용 이력을 관리할 수 있는 서비스를 제공한다.[9] 또한, 헬스아바타 사업은 한 개인에 대한 모든 건강 데이터의 총합을 정의하고, 모바일 운영 시스템을 구축하여, 개인 맞춤형 참여의학을 실현한다. 개인 헬스아바타 앱의 개인 건강정보에 안전하게 접근하여 맞춤형 서비스를 제공한다.[10] 이러한 연구들은 표준 CDM 기반 약물감시 시스템 발전에 있어 중요한 기여를 하고 있다.

2.2 디지털 헬스케어 분야에서의 연합학습

성공적인 디지털 헬스케어 서비스 개발을 위해 의료데이터의 필요성이 대두되고 있다. 그러나 국내 의료데이터는 병원마다 데이터 연동의 어려움과 개인정보 보호의 제한 때문에 중앙 집중식 관리가 어렵다. 이 문제를 해결하기 위해 카카오헬스케어는 병원에서의 데이터 반출 문제를 해소하고, 각기 다른 형태의 데이터를 공통 항목으로 활용할 수 있는 연합학습을 이용한 헬스케어 솔루션을 제안하였다.[11]

분산형 데이터 모델인 분산형 바이오헬스 빅데이터 사업단(FEEDER-Net)은 2019년부터 개발되어 왔다. 이 프로젝트는 CDM을 기반으로 하여,



〈그림 2〉 다기관 약물감시 연합학습 라이프사이클 관리 플랫폼 전체 구성도

데이터를 한 곳에 모으는 대신 각 기관에 분석 프로그램 코드를 보내고, 기관 내에서 분석된 통계 결과값만을 취합하여 연구를 수행한다.[12]

전 세계적으로도 연합학습을 활용한 헬스케어 연구를 수행하고 있다. 유럽연합은 '유럽연합 연합학습 기반 신약개발 가속화 프로젝트' EU-MELLODDY를 추진하고 있으며[13,14], 이 프로젝트에는 10개의 제약기업과 7개의 AI 기업이 참여하고 있다. 한국에서는 EU-MELLODDY보다 성능과 실용성이 개선된 AI 신약개발 플랫폼 K-MELLODDY를 구축해[15] 선진국과의 신약개발 격차를 해소하고, 글로벌 경쟁력을 확보하기 위해 마련. 국내 22개 제약기업을 비롯한 다수의 AI 기업과 IT 기업, 대학, 공공기관이 참여하는 전방위적 AI 신약개발 협력사업으로, 직접적 데이터 공유 없이 분산된 데이터를 한 곳으로 집적시키는 효과를 내는 기술을 활용해 우수한 성능의 공용 AI 모델을 개발 연구를 수행하고 있다.

Ⅲ. 연구내용

3.1 연구의 개념도

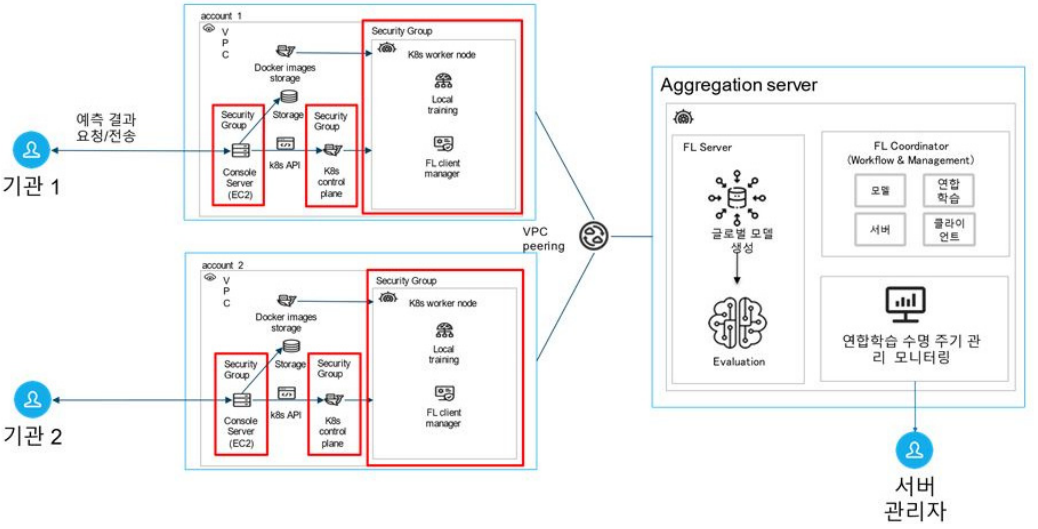
본 연구에서 제안하는 시스템은 그림 1과 같이 고품질 표준 데이터와 프라이버시/개인화 기술을 활용한 연합학습 기반 다기관 약물감시가 가능하다. 각 기관의 데이터는 다른 기관과 공유되지 않으며, 약물감시 파이프라인을 통해 피부 알레르기, 간독성 등과 같은 각 기관 약물 부작용 예측 모델을 생성한다. 각 기관의 로컬 모델들은 퍼블릭 클라우드 환경의 CDM FARM에서 관리되며, 해당 기관의 데이터로 모델을 학습한다. 각 로컬 모델들은 FL 서버와 연결되어 모델의 파라미터를 공유하고 글로벌 모델을 생성 및 업데이트한다. 로컬 모델의 성능에 영향을 미치는 글로벌 모델이 효과적으로 구축되었는지 확인하기 위해, 약물 부작용 예측에 대한 글로벌 모델의 성능을 평가하고 이를 관리한다. 안정적인 약물감시 연합학습 운영을 위해 FL 코디네이터로 연합학습의 전체 프로세스를 관리하고 모니터링한다. 약물 부작용의 결과는 각 기관 어플리케이션에 전달되며 의료진들이 이 정보를 열람하고 결과를 확인할 수 있다.

3.2 다기관 약물감시 연합학습 운영

다기관 간 데이터의 이동없이 약물감시 연합학습을 수행하기 위해, 여러 기관에 약물 부작용 예측을 위한 자동화된 데이터 전처리 및 모델 통합, 배포 기능이 필요하다. 지속적인 약물감시 연합학습 운영을 위해 연합학습 라이프사이클을 관리해야 한다. 다기관 약물감시를 위한 연합학습 라이프사이클 관리 플랫폼을 클라우드 환경에서 운영하여 각 기관의 약물감시에 대한 보안, 관리, 운영, 모니터링 환경을 지원해야 한다. 이러한 기능을 제공하기 위해 그림 2와 같이 약물감시를 위한 연합학습 플랫폼(Pharmacovigilance FedOpsNet)를 제안한다.

3.3 클라우드 기반 표준 CDM FARM

다기관 약물감시 연합학습 시스템의 보안과 확장을 위해 클라우드 환경에서 해당 시스템을 운영해야 한다. 그림 3과 같이 각 기관에 해당되는 VPC(Virtual Private Cloud) 및 서브넷 환경을 구축하여 서로 다른 환경 간의 네트워크 격리를 강화함으로써 데이터 유출을 방지한다. 각 기관과 연합학습 서버 VPC 간에 피어링을 설정하여 안전한 통신을 활성화한다. 이는 연합학습 서버/클라이언트 간 필요한 연결과 프라이버시 모델 전송을 원활하게 한다. 각 기관의 VPC 및 서브넷에 대한 적절한 보안 그룹 및 네트워크 ACL을 구성하여 불필요한 트래픽을 차단하고 필요한 트래픽만을 허용하여 안정성을 유지한다. 각 기관은 로컬 모델에 대한 프라이버시를 유지하며,

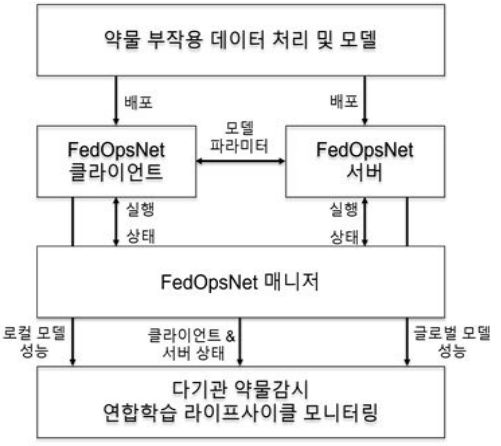


〈그림 3〉 클라우드 기반 연합학습 지원 환경 구성도

로컬 모델 업데이트 및 그래디언트 전송 중에도 외부로 노출되지 않아 데이터 유출의 위험을 최소화한다. 데이터 교환 및 모델 업데이트가 인증된 사용자에 의해 수행되어 보안을 강화한다. 이러한 방안들은 연합학습 시스템의 보안성을 높이고, 참여 기관들의 데이터 프라이버시를 보호하여 약물감시 예측을 수행한다.

3.4 지속적인 관리 기반 약물감시 연합학습

본 연구에서 개발하고자 하는 다기관 약물감시 연합학습 플랫폼은 지속적으로 관리 가능하며, 자동화된 데이터 처리 및 모델 배포를 통해 다양한 장애를 유발하는 특정 약물 부작용 예측 기술을 지속적으로 배포, 통합, 관리한다. 이 플랫폼은 그림 4와 같이 FedOpsNet 클라이언트와 서버를 활용하여 각 기관의 표준 CDM 기반으로 로컬 모델을 학습하고, 이를 서버와 통신하여 글로벌 모델을 생성함으로써 다기관 데이터 공유 없이 약물감시 연합학습을 수행한다. FedOpsNet 매니저를 통해 다기관의 약물감시 연합학습 상황과 서버의 글로벌 모델 생성 과정을 지속적으로 관찰하여 안정적인 연합학습을 지원한다.



〈그림 4〉 지속 가능한 약물감시 연합학습 프로세스

제안하는 연구에서는 다기관 약물감시 연합학습 라이프사이클 관리를 위한 모니터링 시스템을 포함한다. 이 시스템은 각 기관의 연합학습 상황 및 서버의 글로벌 모델 상태를 추적, 관찰 및 관리하는 대쉬보드를 제공함으로써, 연합학습 모델의 지속적인 관리와 자동화된 연합학습 작업 흐름의 효율적인 운영을 지원한다. 이러한 통합적인 접근은 다기관 약물감시 연합학습의 안정성을 지속적으로 향상시키는 데 기여한다.

Ⅳ. 결론

본 연구에서 제안하는 다기관 약물감시 연합학습 플랫폼은 디지털 헬스케어 시장의 성장과 의료 데이터의 중요성을 고려하여 설계되었다. 이 플랫폼은 의료 데이터의 안전한 활용, 효율적인 약물 부작용 예측 및 관리를 가능하게 하는 혁신적인 접근 방식을 제시한다. 특히, 본 연구는 데이터의 중앙집적화 없이 각 기관의 자체 데이터를 활용하여 연합학습을 수행함으로써, 개인정보 보호와 데이터 보안을 강화하는 동시에, 고도화된 약물감시 시스템을 제안하였다.

데이터 보안과 프라이버시 보호를 위한 방법으로, 다기관 VPC및 서브넷 구성을 통해 각 기관 간 네트워크 격리를 강화하였고, FedOpsNet 클라이언트와 서버를 통해 안전한 통신 및 글로벌 모델 생성이 가능하다. 또한, 연합학습 환경에서의 모델 전송 및 저장에 대한 암호화 강화, 각 기관의 프라이버시 유지 및 통신을 보호하기 위한 인증 및 인가 메커니즘 구현 등을 통해 시스템의 보안성을 높였다.

본 연구의 핵심적인 부분은 지속적으로 관리 가능한 연합학습 플랫폼의 구현이다. 자동화된 데이터 처리 및 모델 배포, FedOpsNet 매니저를 통한 플랫폼 관리, 그리고 다기관 약물감시 연합학습 라이프사이클을 위한 모니터링 시스템을 통해, 플랫폼의 안정적인 운영과 효율적인 관리가 가능하다. 이는 다기관 데이터를 효과적으로 활용하여 정확한 약물 부작용 예측과 관리를 가능하게 함으로써, 의료 서비스의 질을 향상시키고 환자 안전을 보장하는 데 기여한다.

이러한 연합학습 플랫폼의 개발과 구현은 디지털 헬스케어 시장의 빠른 성장과 의료 데이터의 중요성에 부응하는 것이며, 향후 신약 개발과 같은 더 넓은 범위의 의료 분야에도 중요한 영향을 미칠 것으로 기대된다. 본 연구가 제시한 플랫폼과 방법론은 연합학습을 기반으로 한 디지털 헬스케어 솔루션의 발전에 중요한 발판을 마련하였으며, 향후 관련 분야의 연구 및 개발에 있어 중요한 기준점을 제공할 것으로 예상된다.

참 고 문 헌

[1] Jeong Mee Kim, Young Mee Lee, “Trends of Adverse Drug Reactions (ADRs)-Related Admissions and Hospital Acquired ADRs in a Korean Tertiary Hospital”, J. Kor. Soc. Health-Syst. Pharm. 32(4), pp.318-334, 2015.

[2] 강민규, 강동윤, 강혜련, 이수현, 김주한, “약물유해반응 관리의 새 흐름: 수동적 유해사례 수집에서 분산형 약물감시망을 이용한 대규모 능동적 검색으로의 전환”, 약물역학위해관리학회지 9(1), pp.1-10, 2017.

[3] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, “Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis,” IEEE Journal of Biomedical and Health Informatics 22(5), pp.1589-1604, 2017.

[4] Si Y, Weng C, “An OMOP CDM-Based Relational Database of Clinical Research Eligibility Criteria”, Stud. Health Technol. Inform. 245, pp.950-954, 2017.

[5] S. Warnat-Herresthal, H. Schultze, K. L. Shastry, S. Manamohan, S. Mukherjee, V. Garg, R. Sarveswara, K. Händler, P. Pickkers, and A. N. Ahmad, “Swarm learning for decentralized and confidential clinical machine learning,” Nature, 594(7862), pp. 265-270, 2021.

[6] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen, and S. Bakas, “Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data,”

Scientific Reports 10(1), pp.1-12, 2020.
[7] Mohemmed Sha M and Mohamudha Parveen Rahamathulla, “Cloud-based Healthcare data management Framework”, KSII Transactions on Internet and Information Systems 14(3), pp.1014-1025, 2020.
[8] 김주한, “다기관 빅데이터 기반 약물 부작용 감시체계 및 지식베이스 구축 사업”, 보건의료 빅데이터 R&D, 2018.
[9] 건강보험심사평가원, “내가 먹는 약! 한눈에”, https://www.hira.or.kr/rb/dur/form.do?pgmid=HIRAA050300000100
[10] Kim JH, “Health avatar: An informatics platform for personal and private big data”, Health Inform. Res. 20(1), pp.1-2, 2014.
[11] 카카오헬스케어, “카카오가 만드는 ‘RWD’, 파일럿 딱지 떼고 전력 질주하나”, http://www.hitnews.co.kr/news/articleView.html?idxno=47845
[12] G.H. Lee, J. Park, J. Kim, Y. Kim, B. Choi, R.W. Park, S.Y. Rhee, S.-Y. Shin, “Feasibility Study of Federated Learning on the Distributed Research Network of OMOP Common Data Model,” Healthcare Informatics Research 29(2), pp.168-173, 2023.
[13] W. Heyndrickx, “MELLODDY: Cross-pharma Federated Learning at Unprecedented Scale Unlocks Benefits in QSAR without Compromising Proprietary Information,” Journal of Chemical Information and Modeling, Article ASAP, 2023.
[14] W. Aziz, M. Vančo, “MELLODDY: Federated machine learning for drug discovery on AWS,” AWS re:Invent, Nov. 29-Dec. 3, 2021.
[15] K-MELODDYY, “K-멜로디 프로젝트가 뭐길래 제약바이오협회가 나설까?”, https://www.doctorsnews.co.kr/news/articleView.html?idxno=150807

저 자 소 개

양 세 모(SeMo Yang)
· 2022년: 가천대학교 컴퓨터공학과 (공학사)
· 2022년~현재: 가천대학교 컴퓨터공학과(석사과정)
<관심분야> 인공지능, IoT, 빅데이터 분석, 디지털 트윈

송 인 서(InSeo Song)
· 2020년~현재: 가천대학교 컴퓨터공학과(학사과정)
<관심분야> 인공지능, IoT, 빅데이터 분석, 디지털 트윈

이 강 윤(KangYoon Lee)
· 1986년: 연세대학교 전자공학과 (공학사)
· 1996년: 연세대학교 전자계산학과 (공학석사)
· 2010년: 숭실대학교 IT정책경영(공학박사)
· 2016년~현재: 가천대학교 컴퓨터공학과 교수
<관심분야> 인공지능, IoT, 빅데이터 활용, 솔루션, 디지털 트윈