FL@META

FL@META, https://github.com/Kwangkee/FL/blob/main/FL%40Meta.md

Where to Begin? On the Impact of Pre-Training and Initialization in Federated Learning, https://arxiv.org/abs/2210.08090

연합 학습의 종종 인용되는 과제는 이질성의 존재이다.
- 데이터 이질성은 서로 다른 클라이언트의 데이터가 매우 다른 분포를 따를 수 있다는 사실을 의미합니다.
- 시스템 이질성은 클라이언트 장치의 시스템 기능이 다르다는 사실을 의미합니다.
상당수의 연합 최적화 방법이 이 문제를 해결한다.

문헌에서 경험적 평가는 일반적으로 무작위 초기화부터 연합 훈련을 시작한다. 그러나 연합 학습의 많은 실제 응용 프로그램에서 서버는 연합 훈련을 시작하기 전에 모델을 사전 교육하는 데 사용할 수 있는 훈련 작업에 대한 프록시 데이터에 액세스할 수 있다.
우리는 네 가지 표준 연합 학습 벤치마크 데이터 세트를 사용하여 연합 학습에서 사전 훈련된 모델에서 시작하는 것의 영향을 경험적으로 연구한다.

당연히 사전 훈련된 모델에서 시작하면 목표 오류율에 도달하는 데 필요한 교육 시간이 단축되고 무작위 초기화를 시작할 때보다 더 정확한 모델(최대 40\%)의 교육이 가능하다.
놀랍게도, 우리는 또한 사전 훈련된 초기화로부터 연합 학습을 시작하는 것이 데이터와 시스템 이질성의 영향을 모두 감소시킨다는 것을 발견했다.

연합 최적화 방법을 제안하고 평가하는 향후 연구는 무작위 및 사전 훈련된 초기화부터 시작할 때 성능을 평가하는 것을 권장한다. 우리는 또한 이 연구가 연합 최적화에서 이질성의 역할을 이해하는 데 대한 추가 연구를 위한 몇 가지 질문을 제기한다고 믿는다.

FL@META

FL@META, https://github.com/Kwangkee/FL/blob/main/FL%40Meta.md

Where to Begin? On the Impact of Pre-Training and Initialization in Federated Learning, https://arxiv.org/abs/2210.08090

In the literature, empirical evaluations usually start federated training from random initialization. However, in many practical applications of federated learning, the server has access to proxy data for the training task that can be used to pre-train a model before starting federated training.
We empirically study the impact of starting from a pre-trained model in federated learning using four standard federated learning benchmark datasets.

Unsurprisingly, starting from a pre-trained model reduces the training time required to reach a target error rate and enables the training of more accurate models (up to 40₩%) than is possible when starting from random initialization.
Surprisingly, we also find that starting federated learning from a pre-trained initialization reduces the effect of both data and system heterogeneity.

We recommend that future work proposing and evaluating federated optimization methods evaluate the performance when starting from random and pre-trained initializations. We also believe this study raises several questions for further work on understanding the role of heterogeneity in federated optimization.

A.2 IMPLEMENTATION DETAILS We implemented all algorithms in Pytorch (Paszke et al., 2017) and evaluated them on a cluster of machines, each with eight NVidia V100 GPUs. We evaluate our experiments in FLSim 4
. 4 https://github.com/facebookresearch/FLSim

Client Selection

Empirical Measurement of Client Contribution for Federated Learning with Data Size Diversification, https://ieeexplore.ieee.org/document/9906094

Client contribution evaluation is crucial in federated learning(FL) to effectively select influential clients. Contrary to data valuation in centralized settings, client contribution evaluation in FL faces a lack of data accessibility and consequently challenges stable quantification of the impact of data heterogeneity. To address this instability of client contribution evaluation, we introduce an empirical method, Federated Client Contribution Evaluation through Accuracy Approximation(FedCCEA), which exploits data size as a tool for client contribution evaluation.

클라이언트 기여 평가는 영향력 있는 클라이언트를 효과적으로 선택하기 위해 연합 학습 (FL) 에서 중요하다. 중앙 집중식 설정의 데이터 평가와 달리 , FL 의 클라이언트 기여 평가는 데이터 접근성의 부족에 직면하여 결과적으로 데이터 이질성의 영향을 안정적으로 정량화하는 데 어려움을 겪는다. 이러한 클라이언트 기여 평가의 불안정성을 해결하기 위해 데이터 크기를 클라이언트 기여 평가 도구로 활용하는 경험적 방법인 FedCCEA (FedCCEA) 를 소개한다.

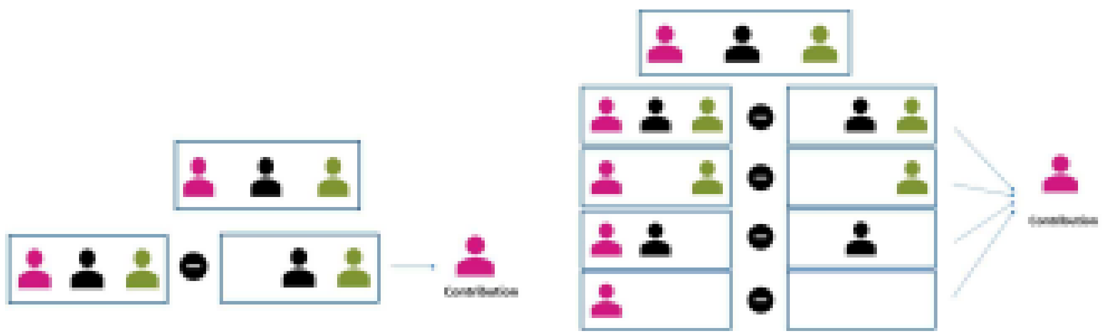
After several FL simulations, FedCCEAapproximates the test accuracy using the sampled data size and extracts the client contribution from the trained accuracy approximator. In addition, FedCCEAgrants data size diversification , which reduces the massive variation in accuracy resulting from game-theoretic strategies. Several experiments have shown that FedCCEAstrengthens the robustness to diverse heterogeneous data environments and the practicality of partial participation.

Client Selection

[NIPA] 인공지능중심 산업융합 집적단지 조성사업 연구개발 - 헬스케어 AI 융합 연구개발,
https://appliedai.skku.edu/appliedailab/ongoing_prj.do?mode=view&articleNo=17788&article.offset=0&articleLimit=10

노이즈가 가미된 연합학습 환경에 대한 클라이언트 기여도 측정 방법의 적합성 평가,
https://appliedai.skku.edu/appliedailab/domestic_pub.do?mode=view&articleNo=25863&article.offset=0&articleLimit=10

연합학습은 분산된 환경에서 직접 데이터를 접근하지 않고 각 클라이언트에서 학습한 모델 파라미터를 통합하여 연합 모델을 생성시키는 분산 머신러닝 기술이다. 연합 모델 성능 향상을 위해 연합학습 통합 알고리즘에 대한 연구가 활발하게 진행되고 있는 반면, 클라이언트 기여도 측정 방법 및 클라이언트 제거 기술에 대한 연구도 하나의 연합학습 연구 분야로 급부상하고 있다. 특히, 노이즈가 투입되어 데이터가 훼손될 수 있는 환경에서 '훼손된 클라이언트'(corrupted clients)를 클라이언트 기여도로 선별하는 기술이 필요하다. 본 논문에서는 연합학습에서 클라이언트 기여도 측정으로 기존에 연구된 대표적인 두 가지 방법, Federated LOO와 Federated SV를 소개한다. 이후 이 두 방법이 노이즈가 가미된 연합학습 환경에서 적절하게 작동되는지 노이즈가 투입된 환경에서 실험을 통해 적합성을 평가한다.



(a) Federated LOO (b) Federated SV

그림 1. 클라이언트 기여도 측정 방법 설명

FL @ Medical/Healthcare

2022_Fall

x

FL @ Medical/Healthcare

Reliability and Performance Assessment of Federated Learning on Clinical Benchmark Data, https://arxiv.org/abs/2005.11756
Federated Learning on Clinical Benchmark Data: Performance Assessment, https://www.jmir.org/2020/10/e20891

삼성병원: 신수용 교수
https://sooyongshin.wordpress.com/2020/11/22/federated-learning/

이제 유명해져서 아는 사람은 다 아는 federated learning..
특히 개인정보 보호가 큰 이슈가 되는 헬스케어 데이터 분석에서 중요한 기술이라고 말이 많은 federated learning..
(*개인적으로 몇년 전부터 이런 기술을 도입해서 사용해야 한다고 강력히 주장하기도 했고.. 특히 가명처리의 대안으로 적극 주장*)
근/데/ 논문이든 뭐든 말만 하지 실제로 하는 사람은 의외로 많지 않은 federated learning.. (*국외 포함. 나도 이 논문 전에는 말만 하던 사람*)
솔직히 말만 많지, 하는 사람이 많이 없어서 그냥 시작했다..

목적은
1.실제 헬스케어 데이터에서 동작하는지 검증
2.실제 network 환경에서 동작하는지 검증

의료 데이터에 쓰면 좋다는 논문은 진짜 많은데..
실제로 검증한 논문은 거의 없다.

FL @ Medical/Healthcare

[CVPR 2022, NVIDIA] Closing the Generalization Gap of Cross-Silo Federated Medical Image Segmentation, https://openaccess.thecvf.com/content/CVPR2022/html/Xu_Closing_the_Generalization_Gap_of_Cross-Silo_Federated_Medical_Image_Segmentation_CVPR_2022_paper.html
Github: https://github.com/NVIDIA/NVFlare/examples/FedSM
Github: https://github.com/NVIDIA/NVFlare/tree/dev/examples

교차 사일로 연합 학습(FL)은 불충분한 데이터, 데이터 개인 정보 보호 및 훈련 효율성이라는 중요한 문제를 해결할 수 있기 때문에 최근 딥 러닝을 통한 의료 영상 분석에서 많은 관심을 끌고 있다.

그러나 FL에서 훈련된 모델과 중앙 집중식 교육에서 훈련된 모델 사이에는 일반화 격차가 있을 수 있다. 이 중요한 문제는 참여하는 클라이언트의 로컬 데이터의 비-iid 데이터 분포에서 발생하며 클라이언트 드리프트로 잘 알려져 있다.

본 연구에서는 처음으로 의료 영상 분할 작업에 대한 중앙 집중식 훈련과 비교하여 클라이언트 드리프트 문제를 피하고 일반화 격차를 성공적으로 좁히기 위한 새로운 훈련 프레임워크 FedSM을 제안한다. 우리는 또한 제안된 프레임워크 FedSM에서 이를 해결하기 위한 새로운 개인화된 FL 목표 공식과 SoftPull 방법을 제안한다.

우리는 비볼록 매끄러운 목적 함수를 최적화하기 위한 수렴을 보장하기 위해 엄격한 이론적 분석을 수행한다. 심층 FL을 사용한 실제 의료 이미지 분할 실험은 우리가 제안한 방법의 동기와 효과를 검증한다.

FL @ Medical/Healthcare

[CVPR 2022, NVIDIA] Closing the Generalization Gap of Cross-Silo Federated Medical Image Segmentation, https://openaccess.thecvf.com/content/CVPR2022/html/Xu_Closing_the_Generalization_Gap_of_Cross-Silo_Federated_Medical_Image_Segmentation_CVPR_2022_paper.html
Github: https://github.com/NVIDIA/NVFlare/examples/FedSM
Github: https://github.com/NVIDIA/NVFlare/tree/dev/examples

Cross-silo federated learning (FL) has attracted much attention in medical imaging analysis with deep learning in recent years as it can resolve the critical issues of insufficient data, data privacy, and training efficiency.

However, there can be a generalization gap between the model trained from FL and the one from centralized training. This important issue comes from the non-iid data distribution of the local data in the participating clients and is well-known as client drift.

In this work, we propose a novel training framework FedSM to avoid the client drift issue and successfully close the generalization gap compared with the centralized training for medical image segmentation tasks for the first time. We also propose a novel personalized FL objective formulation and a new method SoftPull to solve it in our proposed framework FedSM.

We conduct rigorous theoretical analysis to guarantee its convergence for optimizing the non-convex smooth objective function. Real-world medical image segmentation experiments using deep FL validate the motivations and effectiveness of our proposed method.

FL @ Medical/Healthcare

Industry-Scale Orchestrated Federated Learning for Drug Discovery, https://arxiv.org/abs/2210.08871
Github: https://github.com/melloddy

연합 학습을 약물 발견에 적용하기 위해 10개의 제약 회사, 학술 연구실, 대형 산업 회사 및 스타트업으로 구성된 유럽 혁신 의약품 이니셔티브(IMI) 프로젝트 MELLODY(grant n°831472)의 맥락에서 새로운 플랫폼을 개발했다. 우리가 아는 한, MELODDY 플랫폼은 개별 파트너의 기밀 데이터 세트를 공유하지 않고도 약물 발견을 위한 글로벌 연합 모델을 만들 수 있는 최초의 산업 규모 플랫폼이었다.

연합 모델은 각 교육 반복에 따라 암호화적이고 안전한 방식으로 기여하는 모든 파트너의 그레이디언트를 집계하여 플랫폼에서 교육되었다. 이 플랫폼은 개인 서브넷에서 Kubernetes 클러스터를 실행하는 AWS(Amazon Web Services) 다중 계정 아키텍처에 배포되었습니다. 조직적으로 서로 다른 파트너의 역할은 플랫폼에서 서로 다른 권리와 권한으로 코드화되어 분산 방식으로 관리되었습니다. 멜로드디 플랫폼은 동반 논문에서 설명하는 새로운 과학적 발견을 창출했다.

FL @ Medical/Healthcare

Industry-Scale Orchestrated Federated Learning for Drug Discovery, https://arxiv.org/abs/2210.08871
https://github.com/melloddy

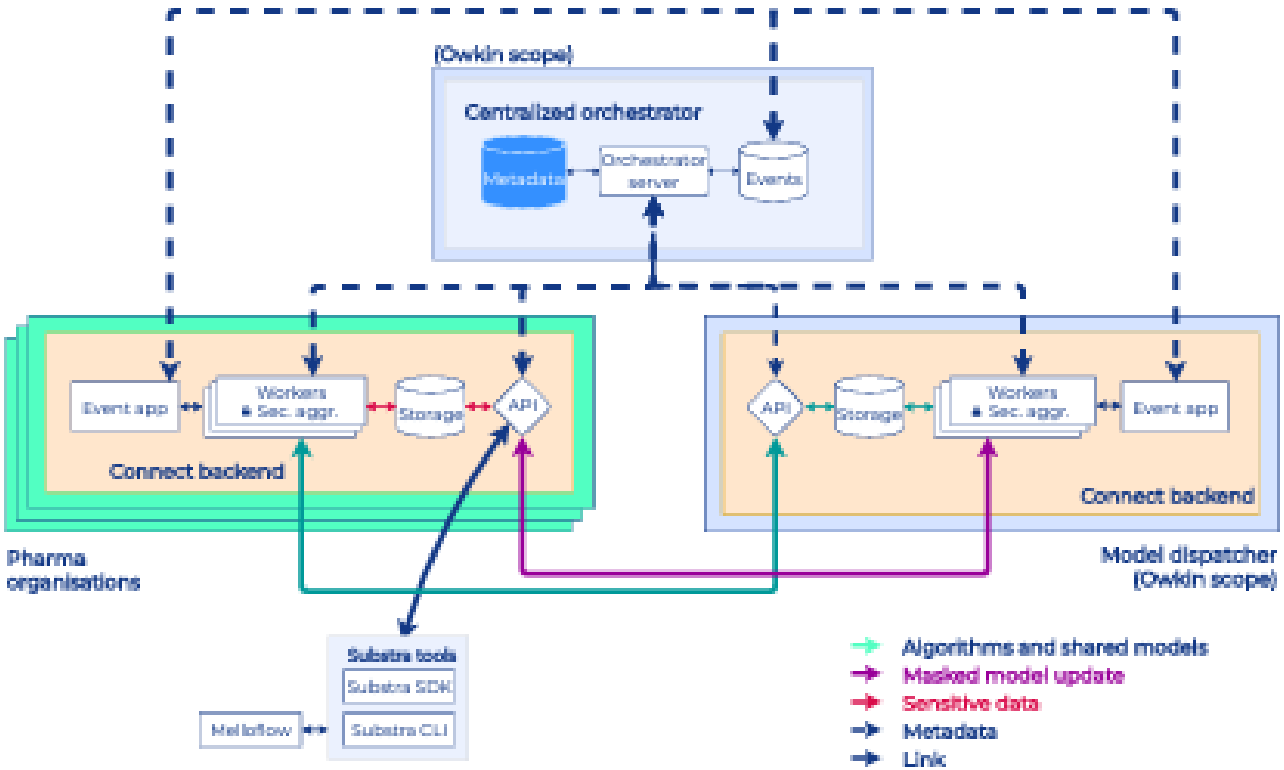


Figure 3: Overview of the deployed platform

FL @ Medical/Healthcare

Reward Systems for Trustworthy Medical Federated Learning, https://arxiv.org/abs/2205.00470
Github: https://github.com/kpandl/Reward-System-for-Trustworthy-Medical-Federated-Learning

연합 학습(FL)은 의료용 기계 학습(ML) 모델을 훈련하기 위해 연구자와 실무자로부터 높은 관심을 받았다. 이러한 모델의 신뢰성을 보장하는 것은 필수적이다. 특히 다른 하위 그룹에 걸친 모델의 예측 성능의 차이로 정의되는 편향은 특정 하위 그룹에 대한 불공정성을 유발할 수 있으며, 이는 신뢰할 수 있는 ML 모델에 바람직하지 않은 현상이다.

본 연구에서는 의료 FL에서 편견이 어느 정도 발생하는지, 보상 시스템을 통해 과도한 편향을 방지하는 방법에 대한 질문을 다룬다.
- 먼저 섀플리 값 근사 방법을 사용하여 교차 사일로 의료 FL의 예측 성과 및 편향에 대한 기관의 기여를 측정하는 방법을 평가한다.
- 두 번째 단계에서는 높은 예측 성능 또는 낮은 편향에 대한 기여를 장려하는 다양한 보상 시스템을 설계한다.
- 그런 다음 우리는 두 가지 모두에 대한 기여를 장려하는 결합 보상 시스템을 제안한다. 환자의 성별과 연령에 따라 정의된 환자 하위 그룹에 초점을 맞춘 여러 의료 흉부 X선 데이터 세트를 사용하여 작업을 평가한다.

우리의 결과는 우리가 편향에 대한 기여를 성공적으로 측정할 수 있다는 것을 보여주며, 통합 보상 시스템은 낮은 편향으로 잘 수행되는 모델에 대한 기여를 성공적으로 장려한다. 스캔의 분할은 전체 편향에 약간만 영향을 미치지만, 주로 한 하위 그룹의 데이터를 가진 기관은 이 하위 그룹에 유리한 편향을 도입한다. 우리의 결과는 예측 성능에만 초점을 맞춘 보상 시스템이 환자에 대한 모델 편견을 제도적 수준으로 전달할 수 있음을 나타낸다. 우리의 연구는 신뢰할 수 있는 ML에 대한 인센티브가 잘 조정된 FL에 대한 보상 시스템을 설계하는 데 도움이 된다.

FL @ Medical/Healthcare

Reward Systems for Trustworthy Medical Federated Learning, https://arxiv.org/abs/2205.00470
Github: https://github.com/kpandl/Reward-System-for-Trustworthy-Medical-Federated-Learning

Federated learning (FL) has received high interest from researchers and practitioners to train machine learning (ML) models for healthcare. Ensuring the trustworthiness of these models is essential. Especially bias, defined as a disparity in the model's predictive performance across different subgroups, may cause unfairness against specific subgroups, which is an undesired phenomenon for trustworthy ML models.

In this research, we address the question to which extent bias occurs in medical FL and how to prevent excessive bias through reward systems.

- We first evaluate how to measure the contributions of institutions toward predictive performance and bias in cross-silo medical FL with a Shapley value approximation method.
- In a second step, we design different reward systems incentivizing contributions toward high predictive performance or low bias.
- We then propose a combined reward system that incentivizes contributions toward both. We evaluate our work using multiple medical chest X-ray datasets focusing on patient subgroups defined by patient sex and age.

Our results show that we can successfully measure contributions toward bias, and an integrated reward system successfully incentivizes contributions toward a well-performing model with low bias. While the partitioning of scans only slightly influences the overall bias, institutions with data predominantly from one subgroup introduce a favorable bias for this subgroup. Our results indicate that reward systems, which focus on predictive performance only, can transfer model bias against patients to an institutional level. Our work helps researchers and practitioners design reward systems for FL with well-aligned incentives for trustworthy ML.

FL @ Medical/Healthcare

Reward Systems for Trustworthy Medical Federated Learning, https://arxiv.org/abs/2205.00470
Github: https://github.com/kpandl/Reward-System-for-Trustworthy-Medical-Federated-Learning

Federated learning (FL) has received high interest from researchers and practitioners to train machine learning (ML) models for healthcare. Ensuring the trustworthiness of these models is essential. Especially bias, defined as a disparity in the model's predictive performance across different subgroups, may cause unfairness against specific subgroups, which is an undesired phenomenon for trustworthy ML models.

In this research, we address the question to which extent bias occurs in medical FL and how to prevent excessive bias through reward systems.

- We first evaluate how to measure the contributions of institutions toward predictive performance and bias in cross-silo medical FL with a Shapley value approximation method.
- In a second step, we design different reward systems incentivizing contributions toward high predictive performance or low bias.
- We then propose a combined reward system that incentivizes contributions toward both. We evaluate our work using multiple medical chest X-ray datasets focusing on patient subgroups defined by patient sex and age.

Our results show that we can successfully measure contributions toward bias, and an integrated reward system successfully incentivizes contributions toward a well-performing model with low bias. While the partitioning of scans only slightly influences the overall bias, institutions with data predominantly from one subgroup introduce a favorable bias for this subgroup. Our results indicate that reward systems, which focus on predictive performance only, can transfer model bias against patients to an institutional level. Our work helps researchers and practitioners design reward systems for FL with well-aligned incentives for trustworthy ML.

FL @ Medical/Healthcare

서울과기대 박종혁 교수, https://scholar.google.com/citations?hl=en&user=IshTErgAAAAJ&view_op=list_works&sortby=pubdate
- FusionFedBlock: Fusion of Blockchain and Federated Learning to Preserve Privacy in Industry 5.0, https://www.sciencedirect.com/science/article/abs/pii/S1566253522001658
- Federated Learning-based secure Electronic Health Record sharing scheme in Medical Informatics, https://ieeexplore.ieee.org/abstract/document/9774951

의료 사이버 물리 시스템은 임상 연구를 위한 전자 건강 기록 데이터의 이동성을 지원하여 새로운 과학적 발견을 가속화한다. 인공지능은 의료 정보학을 향상시키지만, 현재의 중앙 집중식 데이터 훈련과 안전하지 않은 데이터 저장 관리 기법은 민간 의료 데이터를 허가받지 않은 외국 기관에 노출시킨다.

본 논문에서는 의료 정보학이 환자 데이터 프라이버시를 보존하기 위해 연합 학습 기반 전자 건강 기록 공유 체계를 제안한다. 분산 연합 학습 기반 컨볼루션 신경망 모델은 병원에서 로컬로 데이터를 훈련하고 결과를 개인 행성 간 파일 시스템에 저장한다. 2차 글로벌 모델은 로컬 모델을 사용하여 연구 센터에서 훈련된다. 사설 IPFS는 병원에 로컬로 저장된 모든 의료 데이터를 보호합니다.

이 연구의 참신함은 임상 연구 조직에 유용한 병원 생물의학 데이터를 확보하는 데 있다.
- 블록체인과 스마트 계약을 통해 환자는 데이터를 교환하는 대가로 외부 엔티티와 보상을 협상할 수 있다.
평가 결과는 분산형 CNN 모델이 기존의 중앙 집중형 모델과 유사하게 정확도, 민감도 및 특수성에서 더 나은 성능을 발휘한다는 것을 보여준다.
- 프라이빗 IPFS의 성능은 파일 업로드 및 다운로드 시간을 기준으로 블록체인 기반 IPFS를 초과한다.

이 계획은 생물 의학 연구를 위한 임상 연구 센터와 데이터를 공유하기 위한 안전하고 개인 정보 보호 친화적인 환경을 촉진하는 데 적합하다.

FL @ Medical/Healthcare

- Federated Learning-based secure Electronic Health Record sharing scheme in Medical Informatics, https://ieeexplore.ieee.org/abstract/document/9774951

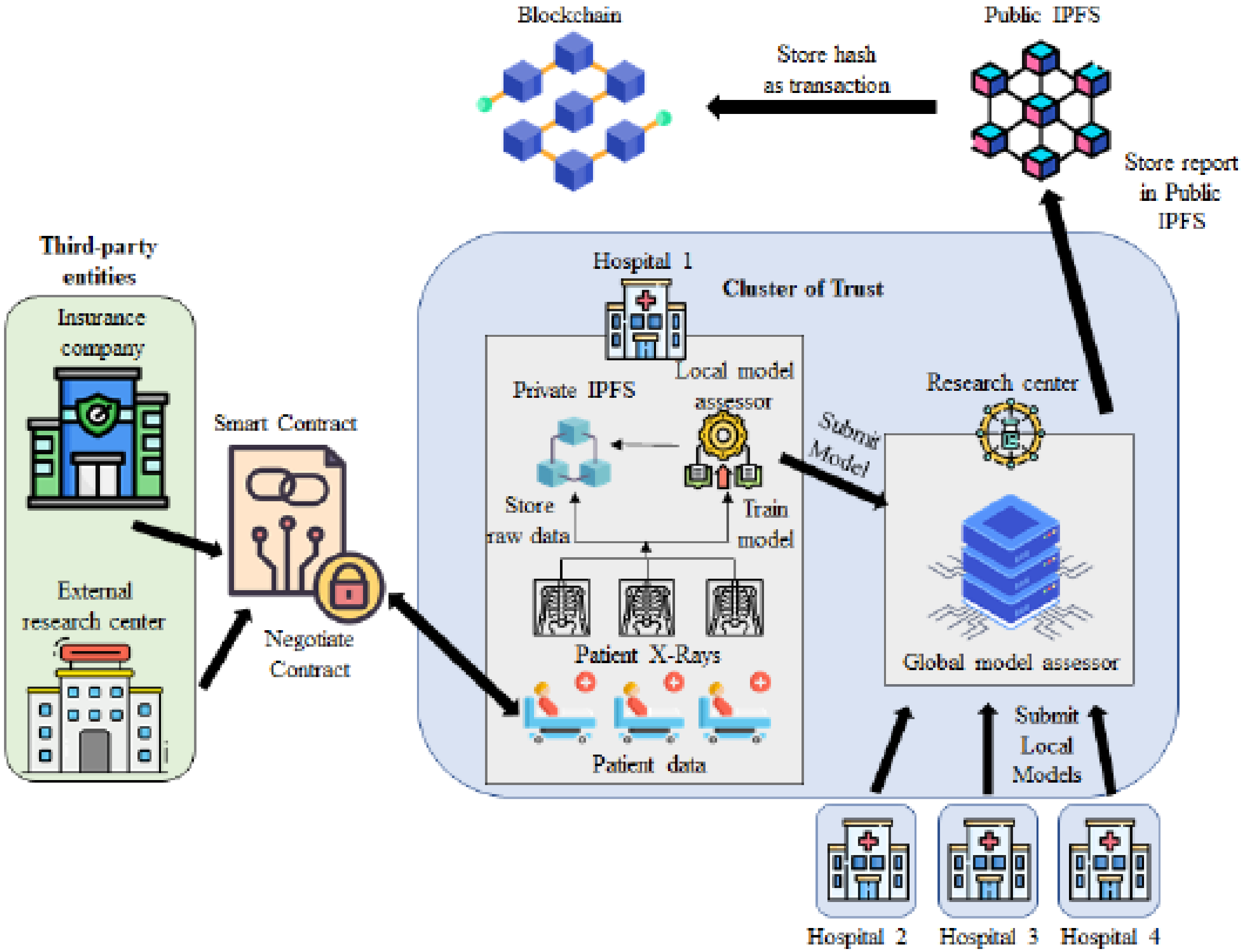


Fig. 1. Secure EHR Sharing Scheme Overview

FL @ Medical/Healthcare

Federated Learning-based secure Electronic Health Record sharing scheme in Medical Informatics, https://ieeexplore.ieee.org/abstract/document/9774951

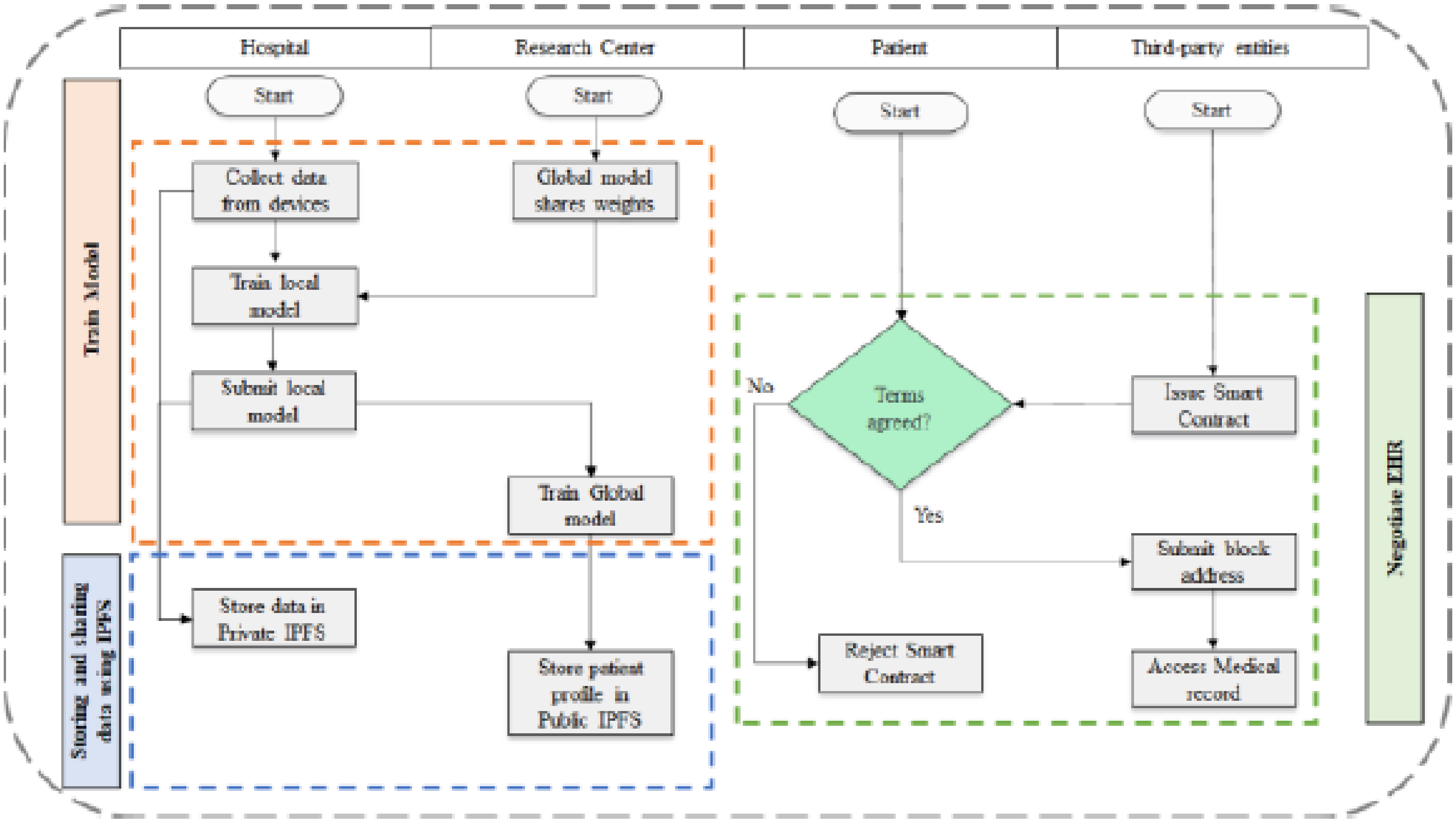


Fig. 2. Secure EHR Scheme process-flow

FL @ Medical/Healthcare

[AAAI 2022] CAreFL: Contribution-Aware Federated Learning for Smart Healthcare, https://ojs.aaai.org/index.php/AAAI/article/view/21505

인공지능(AI)은 의료 산업을 변화시킬 유망한 기술이다. 환자 데이터의 매우 민감한 특성으로 인해 연합 학습(FL)은 종종 스마트 의료 애플리케이션 모델을 구축하는 데 활용된다.

기존의 구축된 FL 프레임워크는 이 분야의 여러 기관에 걸쳐 다양한 데이터 품질과 이기종 데이터 분포의 핵심 문제를 해결할 수 없다.

본 논문에서는 스마트 헬스케어를 위한 기여 인식 연합 학습(CAREFL) 프레임워크를 개발하고 배포한 경험을 보고한다. 효율적이고 개인 정보 보호 방식으로 공정하고 설명 가능한 FL 참여자 기여 평가를 제공하고, 평가 결과를 기반으로 FL 모델 집계 접근법을 최적화한다.

CareFL은 2021년 3월 Yidu Cloud Technology Inc.에 구축한 이후 중국 내 8개 의료기관에서 의료 의사결정 지원 모델을 구축하기 위해 서비스를 제공하고 있습니다. 기존 최선의 접근 방식보다 기여도 평가를 2.84배 빠르게 수행할 수 있으며, 결과 모델의 평균 정확도가 기존 시스템(산업 환경에서 유의미한 수준)보다 2.62% 향상됐다.

우리가 아는 한, 그것은 의료 산업에서 성공적으로 배치된 최초의 기여 인식 연합 학습이다.

FL @ Medical/Healthcare

[AAAI 2022] CAreFL: Contribution-Aware Federated Learning for Smart Healthcare, https://ojs.aaai.org/index.php/AAAI/article/view/21505

It provides fair and explainable FL participant contribution evaluation in an efficient and privacy-preserving manner, and optimizes the FL model aggregation approach based on the evaluation results.

- Since its deployment in Yidu Cloud Technology Inc. in March 2021, CAreFL has served 8 well-established medical institutions in China to build healthcare decision support models.
- It can perform contribution evaluations 2.84 times faster than the best existing approach, and has improved the average accuracy of the resulting models by 2.62% compared to the previous system (which is significant in industrial settings). To our knowledge, it is the first contribution-aware federated learning successfully deployed in the healthcare industry.

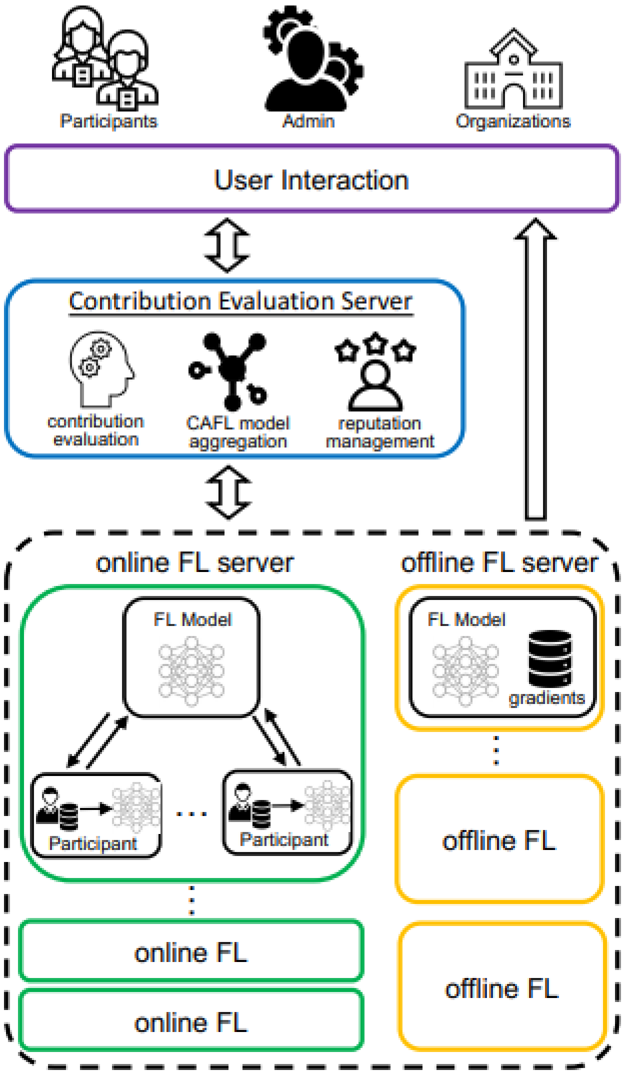


Figure 1: The CAreFL system architecture.

FL @ Medical/Healthcare

[AAAI 2022] CAreFL: Contribution-Aware Federated Learning for Smart Healthcare, https://ojs.aaai.org/index.php/AAAI/article/view/21505

FL participant **contribution evaluation** is an active subfield of FL (Ghorbani and Zou 2019; Jia et al. 2019; Song, Tong, and Wei 2019; Wang et al. 2020; Wei et al. 2020). The aim is to estimate the value of each FL participant by evaluating its impact on the performance of the resulting FL model, without exposing their sensitive local data. To bridge the aforementioned gaps in FL frameworks for smart healthcare, we propose the Contribution-Aware Federated Learning (CAreFL) framework.

The advantages are:

1. **Fast and Accurate Contribution Evaluation**: it is incorporated with our proposed GTG-Shapley (Liu et al. 2022) approach, which can evaluate fair and accurate FL participant contribution in a highly efficient manner.
2. **Contribution-Aware FL Model Aggregation**: during the contribution evaluation process, GTG-Shapley builds a large number of aggregated FL sub-models involving local model updates from different combinations of FL participants. With this knowledge, CAreFL provides a novel FL aggregation approach which selects the best performing sub-model to be distributed to the FL participants for the next round of local training. This differs from FedAvg-based approaches (which always aggregate all received local models), and can better deal with data heterogeneity issues.
3. **Contribution-based FL Participant Reputation Management**: historical contribution evaluation records are converted into reputation values for the FL participants. This information can serve as a basis stakeholder management decision support.

FL @ Medical/Healthcare

[AAAI 2022] CAreFL: Contribution-Aware Federated Learning for Smart Healthcare, https://ojs.aaai.org/index.php/AAAI/article/view/21505

- Hence, the canonical SV cannot be directly used for contribution evaluation in the context of FL.
- The key idea of GTG-Shapley is to opportunistically reduce the need for sub-model retraining with model reconstruction and strategic sampling of combinations of participants. It truncates unnecessary sub-model evaluations to reduce computational costs, while maintaining high accuracy of estimated SVs.

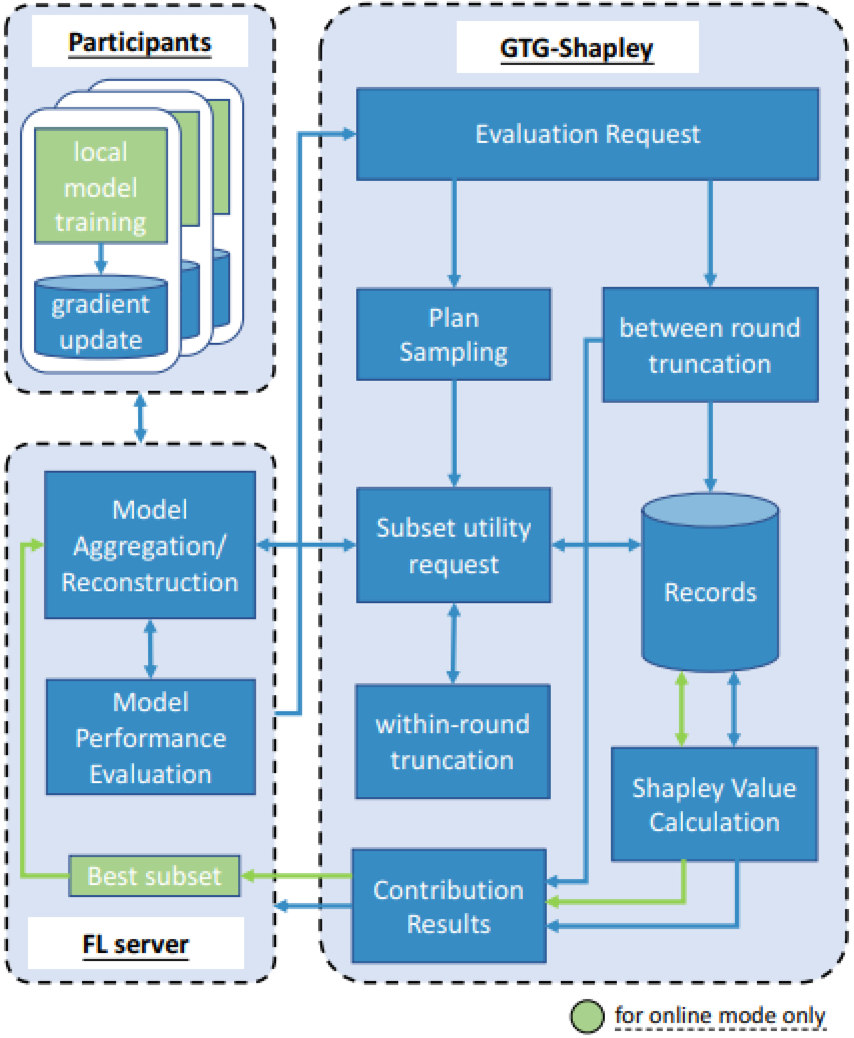


Figure 3: The CAreFL AI Engine

FL @ Medical/Healthcare

[AAAI 2022] CAreFL: Contribution-Aware Federated Learning for Smar
Healthcare, https://ojs.aaai.org/index.php/AAAI/article/view/21505

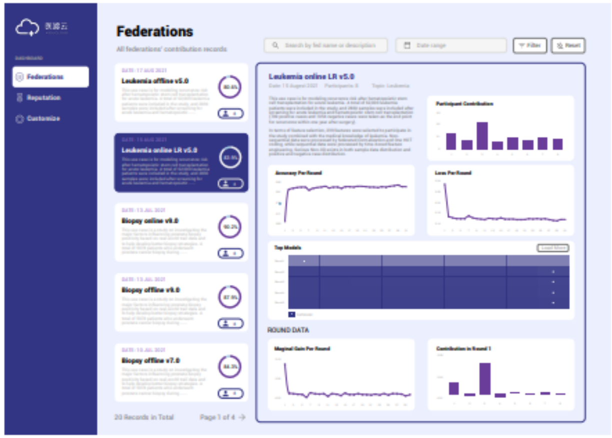


Figure 5: The main user interface of CAreFL.

Application Use and Payoff
The CAreFL framework has been deployed in Yidu Cloud Technology I
lines of their business:
1. clinical research services : Clinical research focuses on training FL m
from multiple hospitals.
2. real-world trial research services. : Real world trial research is often initiated by a pharmaceutical company which aims to leverage data from multiple hospitals to build models.

Both services require data which need to be collected by the hospitals over months or years under their respective Institutional Review Board (IRB) supervision. So far, CAreFL has been used to help eight well-known medical institutions in China to train AI models for risk prediction, disease diagnosis and influence factor analysis.

Conclusions and Future Work
In future, we will continue the explore the applicability of CAreFL in other smart healthcare application scenarios. We will also extend the CAreFL framework with contribution-based data pricing mechanisms (Pei 2020) to support the emergence of an **FL-based healthcare data exchange marketplace**. Eventually, we aim to incorporate these functionalities into the opensource FATE framework and make them available to more developers, researchers and practitioners.

FL @ Medical/Healthcare

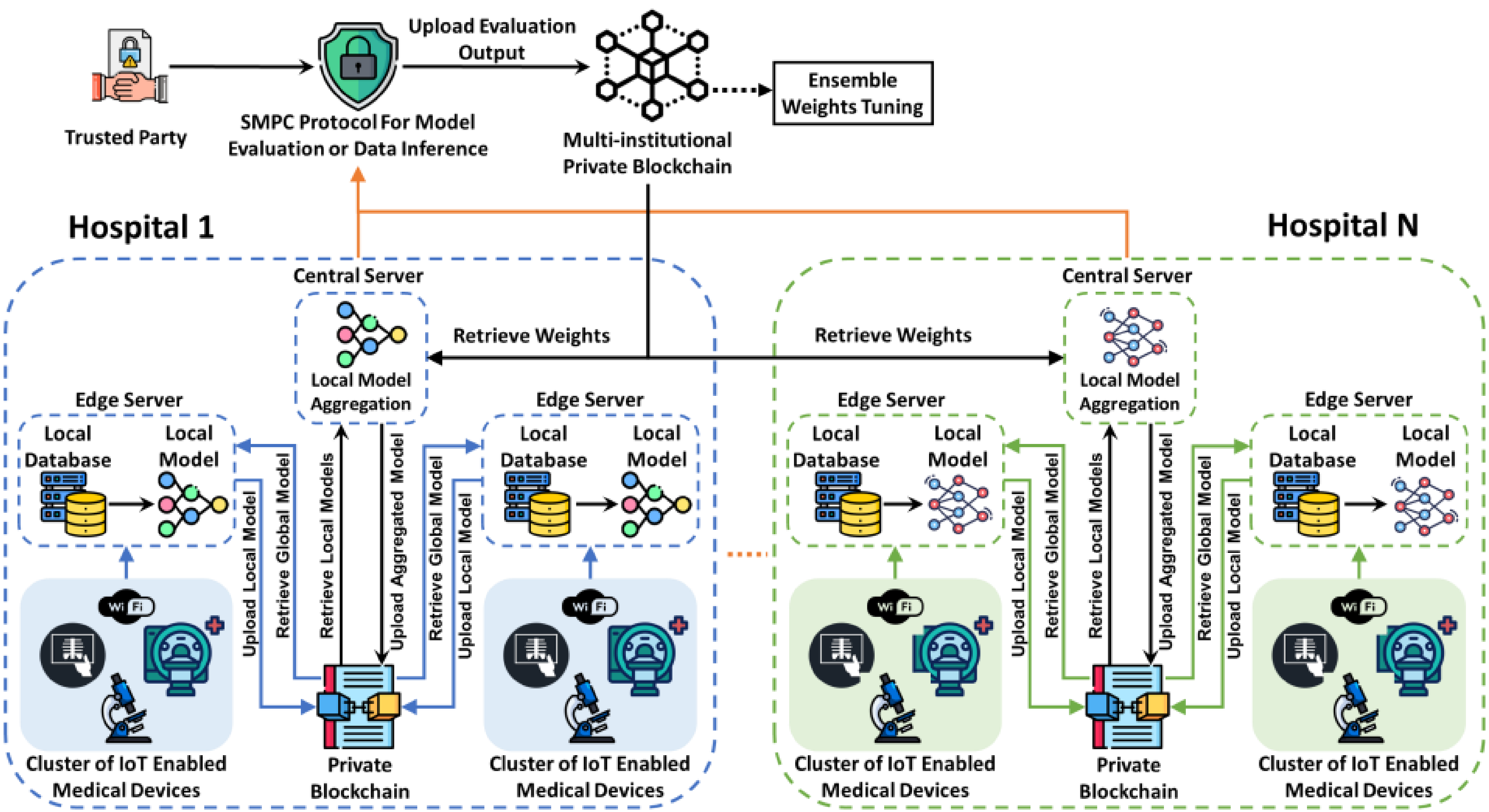
Trustworthy Privacy-preserving Hierarchical Ensemble and Federated Learning in Healthcare 4.0 with Blockchain, https://ieeexplore.ieee.org/abstract/document/9920221?casa_token=3vBmyEn_r1IAAAAA:FL4Xp9bjfO273envprcmHpVJyzkSoa9WQwSD3xtL4fEwIMlXJx-oYeH_QD1HXssChx--vRLRXUfjNA

인터넷과 통신기술(ICT)의 발달로 인더스트리 4.0 시대가 열렸습니다. 이러한 변화에 이어 의료 산업이 Healthcare 4.0이라는 용어를 만들었습니다. Healthcare 4.0에서 IoT 지원 의료 영상 장치를 질병 조기 감지에 사용함으로써 의료 종사자는 의료 기관의 서비스 품질을 높일 수 있습니다. 그러나 Healthcare 4.0은 데이터 개인 정보 보호 문제로 인해 다른 Industry 4.0에 비해 인공 지능 및 빅 데이터에서 여전히 뒤떨어져 있습니다. 또한 기관의 다양한 저장 및 컴퓨팅 기능으로 인해 기관이 동일한 교육 모델 구조를 통합하는 데 제한이 있습니다.

이 논문은 이기종 모델이 사용자의 개인 정보를 침해하지 않고 의료 기관의 데이터에서 협력적으로 학습할 수 있도록 하는 블록체인을 사용한 안전한 다자간 컴퓨팅 기반 앙상블 연합 학습을 제시합니다. 또한 블록체인 속성을 통해 당사자는 중앙 서버에 대한 신뢰 없이 데이터 무결성을 누릴 수 있으며 각 의료 기관에 감사 가능성 및 버전 제어 기능을 제공할 수 있습니다.

FL @ Medical/Healthcare

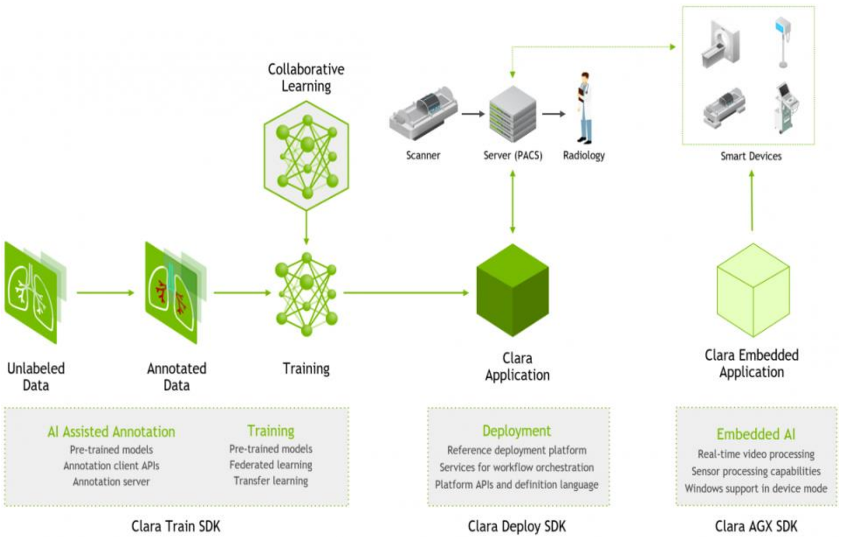
Trustworthy Privacy-preserving Hierarchical Ensemble and Federated Learning in Healthcare 4.0 with Blockchain, https://ieeexplore.ieee.org/abstract/document/9920221?casa_token=3vBmyEn_r1IAAAAA:FL4Xp9bjfO273envprcmHpVJyzkSoa9WQwSD3xtL4fEwIMlXJx-oYeH_QD1HXssChx--vRLRXUfjNA

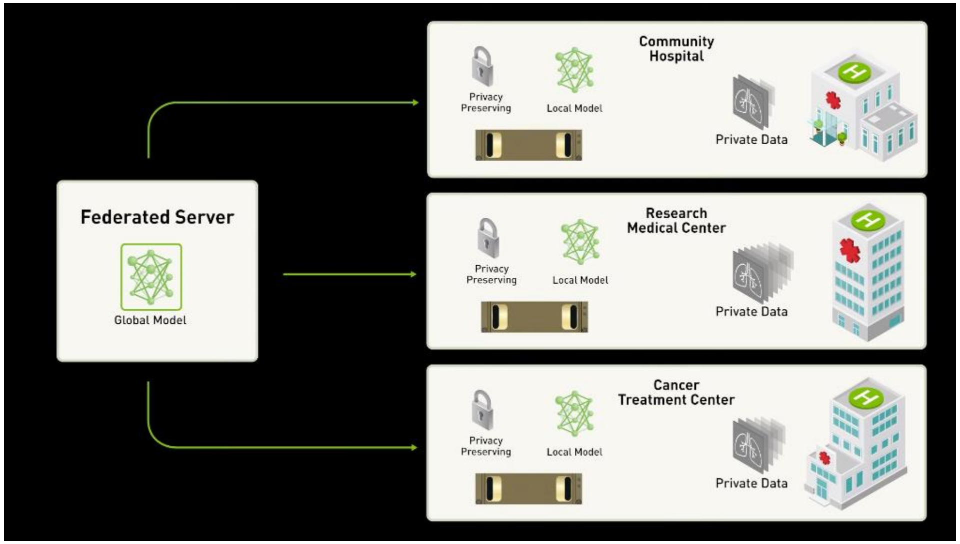


Federated Learning powered by NVIDIA Clara

- An Application Framework Optimized for Healthcare and Life Sciences Developers, https://developer.nvidia.com/clara
- Federated Learning powered by NVIDIA Clara, https://developer.nvidia.com/blog/federated-learning-clara/

- Transforming AI Healthcare with Federated Learning, https://news.developer.nvidia.com/transforming-ai-healthcare-with-federated-learning/
- https://www.nature.com/articles/s41746-020-00323-1





Google Health Studies

Advancing health research with Google Health Studies,
https://blog.google/technology/health/google-health-studies-app/
https://play.google.com/store/apps/details?id=com.google.android.apps.health.research.studies

COVID-19 has highlighted the importance of research in providing information about disease and treatments. However, it's challenging for researchers to recruit enough volunteers so that studies are representative of the general population. To make it easier for leading research institutions to connect with potential study participants, we're introducing the Google Health Studies app with the first study focused on respiratory illness.

Keeping participant data private, safe and secure
Studying respiratory illnesses
We've partnered with researchers from Harvard Medical School and Boston Children's Hospital for the first study, which will help scientists and public health communities better understand respiratory illnesses, including influenza and COVID-19.

- This Respiratory Health Study will be open to adults in the U.S., and will focus on identifying how these types of illnesses evolve in communities and differ across risk factors such as age, and activities such as travel.
- Study participants will use the Google Health Studies app to regularly self-report how they feel, what symptoms they may be experiencing, any preventative measures they've taken, and additional information such as COVID-19 or influenza test results. By taking part in this study, volunteers can represent their community in medical research, and contribute to global efforts to combat the COVID-19 pandemic.

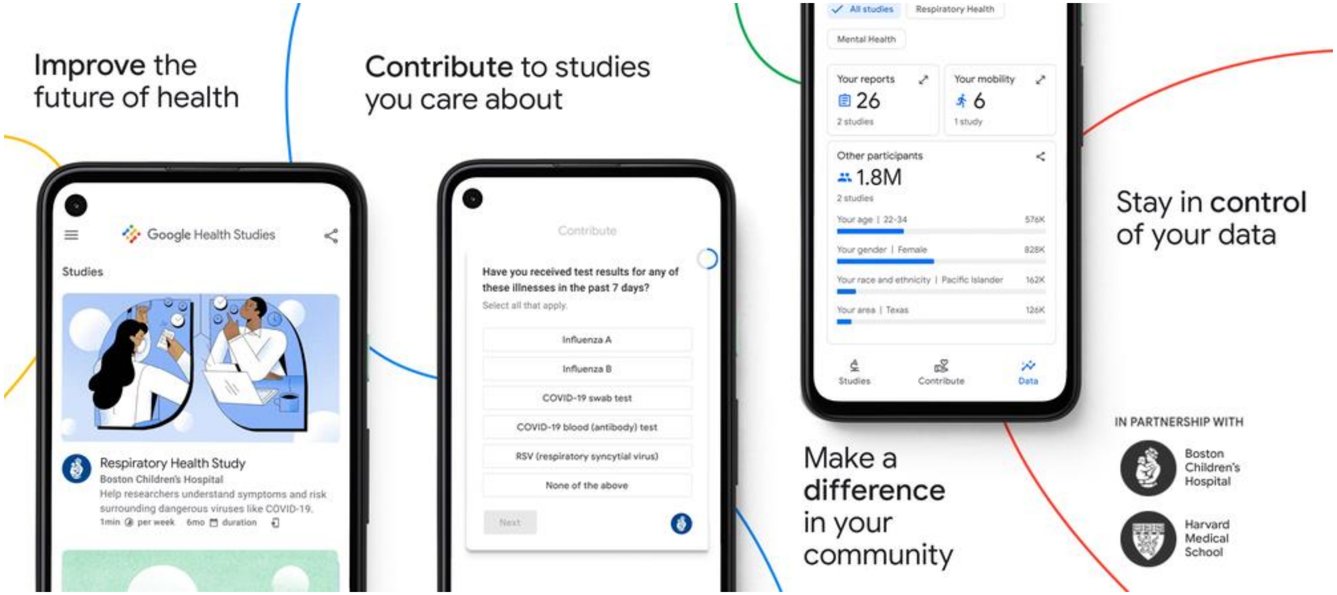
In collaboration with Google Research, this first study utilizes federated learning and analytics—a privacy technology that keeps a person's data stored on the device, while allowing researchers to discover aggregate insights based on encrypted, combined updates from many devices.

Google Health Studies : 연합학습 적용

Blog: https://blog.google/technology/health/google-health-studies-app/
App: https://play.google.com/store/apps/details?id=com.google.android.apps.health.research.studies

The Google Health Studies app is now available in the Google Play Store, and we're inviting people to download the app to join this initial study. We look forward to partnering with health researchers and to making it possible for more people to participate in these important studies.

… this first study utilizes federated learning and analytics—a privacy technology that keeps a person's data stored on the device, while allowing researchers to discover aggregate insights based on encrypted, combined updates from many devices. This means researchers in this study can examine trends to understand the link between mobility (such as the number of daily trips a person makes outside the home) and the spread of COVID-19, This same approach powers typing predictions on Gboard, without Google seeing what individuals type.

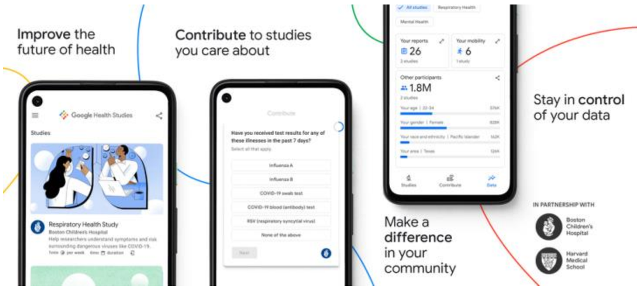


Google Health Studies : 연합학습 적용

Blog: https://blog.google/technology/health/google-health-studies-app/
App: https://play.google.com/store/apps/details?id=com.google.android.apps.health.research.studies

Google Health Studies lets you securely contribute to health research studies with leading institutions, right from your phone. Volunteer for studies that matter to you and represent your community.

Simply download the app and enroll in a study.



Help researchers make advancements in medicine and healthcare:
- Self-report symptoms and other data
- Volunteer for multiple studies in one app
- Track your information with digital health reports
- Learn research findings from the studies you participate in

Help scientists better understand respiratory diseases.
The first study available is a respiratory health study conducted by Boston Children's Hospital and Harvard Medical School. If you participate in this study, you'll provide data to help researchers understand how demographics, health history, behavior, and mobility patterns contribute to the spread of respiratory illnesses. Upcoming studies will research mental health and diabetes.

You're in control of your data: In the respiratory health study, your personal information is kept on your device. Researchers only see aggregated study data combined from all participants. This allows researchers to collect the information needed to advance the study without seeing individual details.

Your input matters: Google Health Studies aims to create opportunities for more people to participate in health research. By contributing, you'll represent your community and start improving the future of health for everyone.

Google Health Studies

- https://health.google/for-everyone/health-studies/
- App: https://play.google.com/store/apps/details?id=com.google.android.apps.health.research.studies
- Blog: https://blog.google/technology/health/google-health-studies-app/

Benefit the public, in private

Protecting your information in the respiratory health study.

Your study data stays on your device

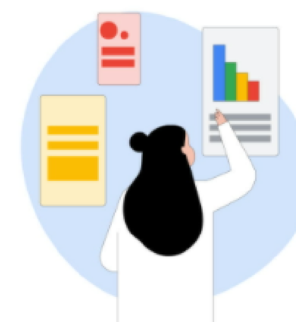
After joining a health study, you'll begin completing weekly surveys. At all times, your individual survey responses, location history and other personally identifiable data stays on your device.

Your device computes statistics based on your study data

During the study, your device receives different queries, computes and summarizes the results based on your individual study data, and encrypts these results for subsequent aggregation with federated analytics.

Participant data gets aggregated

Encrypted summaries from many devices are combined together, using the federated analytics technology. Google and study partners do not receive any individual study data about you.



Research that values your privacy

Combined insights are sent securely to the researchers conducting the study. You can safely contribute to health research knowing your personally identifiable study data will never be available to Google or third parties.

Google Health Studies

Paper: Privacy-first Health Research with Federated Learning, https://www.nature.com/articles/s41746-021-00489-2,
Patent: Privacy-First On-Device Federated Health Modeling and Intervention, https://patents.google.com/patent/US20210090750A1/en

We show—on a diverse set of single and multi-site health studies—that federated models can achieve similar accuracy, precision, and generalizability, and lead to the same interpretation as standard centralized statistical models while achieving considerably stronger privacy protections and without significantly raising computational costs.

At this point, however, only specific large homogenous units of federation, such as at the level of a healthcare system, have been studied in detail in prior work, and the focus has been on traditional classification tasks.

Specifically, health study data is typically non-IID—not independent and identically distributed—which is compounded by the fact that in the federated regime, individual data points are distributed across many devices that participate asynchronously.

This work's primary focus is on cross-device (cross-patient) settings, where the unit of federation is a single individual.

By contrast, in this work, we focus on those scenarios commonly found in epidemiological health studies, specifically studies with many participants, each of whom has relatively small amounts of non-IID, labeled data. The approach described here can be appropriate for health studies involving smartphone/wearable data and virtual clinical studies (also called decentralized clinical studies) that directly recruit individual research participants without relying on clinical sites for recruitment.

Google Health Studies

Paper: Privacy-first Health Research with Federated Learning, https://www.nature.com/articles/s41746-021-00489-2,

Study Topic	Sample Results	Comparison Metric	Traditional Centralized Model[a]	Federated Replications	
				Per-Patient	Per-Silo[b]
Heart Failure	1. Survival Prediction (full model) 2. Survival Prediction (with variable selection)	AUC	0.82 0.82	0.85 0.83	N/A
Diabetes	1. Diabetes prediction at 5-years	AUC	0.84	0.875	N/A
MIMIC-III	1. Inpatient mortality prediction	AUC	0.780± 0.012	0.777 ± 0.011	0.777 ± 0.014
SARS-CoV-2	1. CV2+ve in Female vs. Male 2. CV2+ve in Recent vs. Never Cancer	OR	0.35 (0.32–0.38) 1.88 (1.36–2.60)	0.35 (0.32–0.38) 1.99 (1.45–2.68)	0.35 (0.32–0.38) 2.07 (1.50–2.86)
Avian Influenza	1. Fatality with each day before hospitalization 2. Fatality in Indonesia vs. group of countries	OR	1.33 (1.11–1.60) 0.23 (0.04–1.27)	1.34 (1.12–1.61) 0.25 (0.05–1.37)	1.33 (1.11–1.60) 0.24 (0.04–1.33)
Bacteraemia	1. Relapse with line-associated infection source 2. Relapse with presence of immunosuppression	Coefficient	1.57 (SE: 0.45) 1.07 (SE: 0.41)	1.59 (SE: 0.23) 1.12 (SE: 0.30)	N/A
Azithromycin	1. Adverse events in azithromycin treated	Coefficient	−0.11 (SE: 0.09)	−0.29 (SE: 0.19)	N/A
Tuberculosis	1. Extrapulmonary TB in individuals with HIV	Coefficient	1.16 (SE: 0.09)	1.35 (SE: 0.08)	0.15 (SE: 0.07)[c]

Google Health Studies

Patent: Privacy-First On-Device Federated Health Modeling and Intervention, https://patents.google.com/patent/US20210090750A1/en

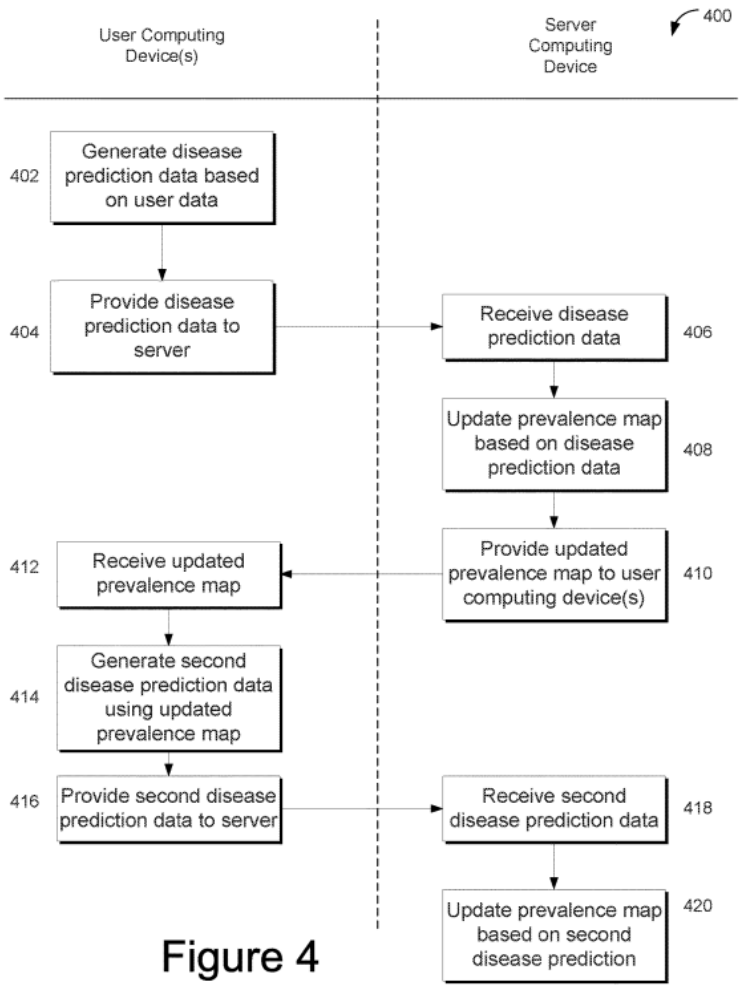


Figure 4

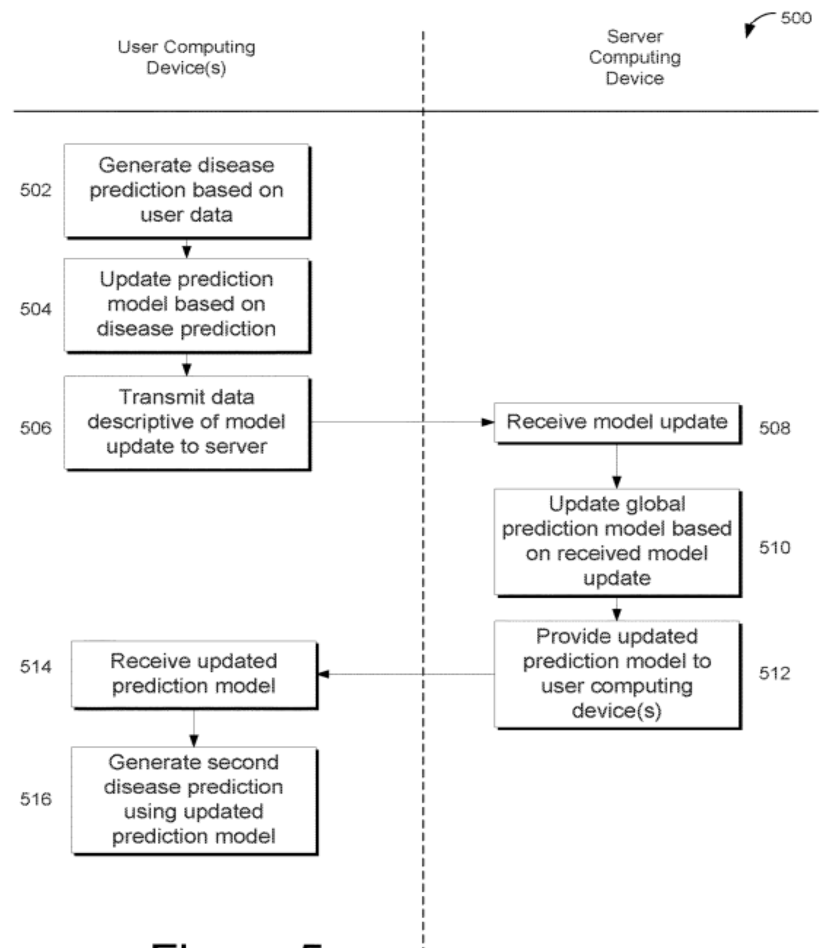


Figure 5

Google Health Studies

Wide Scale Monitoring for Acute Respiratory Infection Using a Mobile-Based Study Platform, https://clinicaltrials.gov/ct2/show/results/NCT04663776

- Sponsor: Boston Children's Hospital
- Collaborator: Google LLC.
- Information provided by (Responsible Party): John Brownstein, Boston Children's Hospital

Brief Summary:
This is a prospective observational study using a mobile study platform (app) that is designed for use on Android phones.

- Study participants will provide baseline demographic and medical information and report symptoms of respiratory infection on a weekly basis using the app.
- Participants will also report use of prevention techniques on the weekly survey.
- Mobility data will be collected passively using the sensors on the participant's smartphone, if the participant has granted the proper device permissions.
- The overall goals of the study are to track spread of coronavirus-like illness (CLI), influenza-like illness (ILI) and non-specific respiratory illness (NSRI) on a near-real time basis and identify specific behaviors associated with an increased or decreased risk of developing these conditions.

Study Population
The study population will be adult Android mobile device users who live within the United States.