Confederated learning: training machine learning models using disconnected data separated by individual, data type and identity for Large-Scale Health System Intelligence

Dianbo Liu, Tim Miller, Ken Mandl
Computational Health Informatics Program, Boston Children's Hospital, Boston, MA
Department of Pediatrics, Harvard Medical School, Boston, MA

This work was supported by CVS Health, by cooperative agreements U01TR002623 from the National Center for Advancing Translational Sciences (NCATS)/National Institutes of Health, and U01HL121518 from the National Heart, Lung, And Blood Institute/National Institutes of Health.

Abstract

A patient's health information is generally fragmented across silos. Though it is technically feasible to unite data for analysis in a manner that underpins a rapid learning healthcare system, privacy concerns and regulatory barriers limit data centralization. Machine learning can be conducted in a federated manner on patient datasets with the same set of variables, but separated across sites of care. But federated learning cannot handle the situation where different data types for a given patient are separated vertically across different organizations and when patient ID matching across different institutions is difficult. We call methods that enable machine learning model training on data separated by two or more degrees "confederated machine learning." We proposed and evaluated a confederated learning to training machine learning model to stratify the risk of several diseases among when data are horizontally separated by individual, vertically separated by data type, and separated by identity without patient ID matching.

Introduction

Significance. Access to a large amount of high quality data is possibly the most important factor for success in advancing medicine with machine learning and data science. However, valuable healthcare data are usually distributed across isolated silos, and there are complex operational and regulatory concerns. Data on patient populations are often *horizontally*

separated, *by individual*, from each other across different practices and health systems. In addition, individual patient data are often vertically separated, *by data type*, across the sites of care, service, and testing. Furthemore, it is often not possible to match patient IDs across different silos in the healthcare system due to operational and privacy issues, which separates patients' data *by identity*.Traditionally, federated machine learning refers to distributed learning on horizontally separated data (Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, Vikas Chandra 2018; Cano, Ignacio, Markus Weimer, Dhruv Mahajan, Carlo Curino, and Giovanni Matteo Fumarola 2016; H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, Blaise Agüera y Arcas 2016). Algorithms are sent to different data silos (sometimes called data nodes) for training. Models obtained are aggregated for inference. Federated learning can reduce data duplication and costs associated with data transfer, while increasing security and shoring up institutional autonomy. (Geyer, R. C., Klein, T., & Nabi, M. 2017; B. K. et al. 2016),(Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, Vikas Chandra 2018; K. J. et al. 2015)(Geyer, R. C., Klein, T., & Nabi, M. 2017; B. K. et al. 2016).

Notably, a patient's vertically separated data may span data types--for example, diagnostic, pharmacy, laboratory, and social services. Machine learning on vertically separated data has used a split neuron network (Praneeth et al. 2018) and homomorphic encryption (Praneeth et al. 2018; Stephen et al. 2017). However, these new methods require patient ID matching ,information communication at each computational cycle and state-of-art computational resource organization, which are usually impractical in many healthcare systems where support for data analysis is not the first priority, high speed synchronized computation resources are often not available, and data availability is inconsistent.

To accelerate a scalable and collaborative rapid learning health system (Friedman, Wong, and Blumenthal 2010; Mandl et al. 2014), we propose a confederated machine learning method that trains machine learning models on data separated by individual, data type and identity using a 3-step approach from data distributed across silos (Qi, Huang, and Peng 2017; Zhang and Xiao 2015; Zhai, Peng, and Xiao 2014). Our confederated learning method does not require individual ID matching, frequent information exchange at each training epoch nor state-of-the-art distributed computing infrastructures. As such, it should be readily implementable, using existing health information infrastructure.

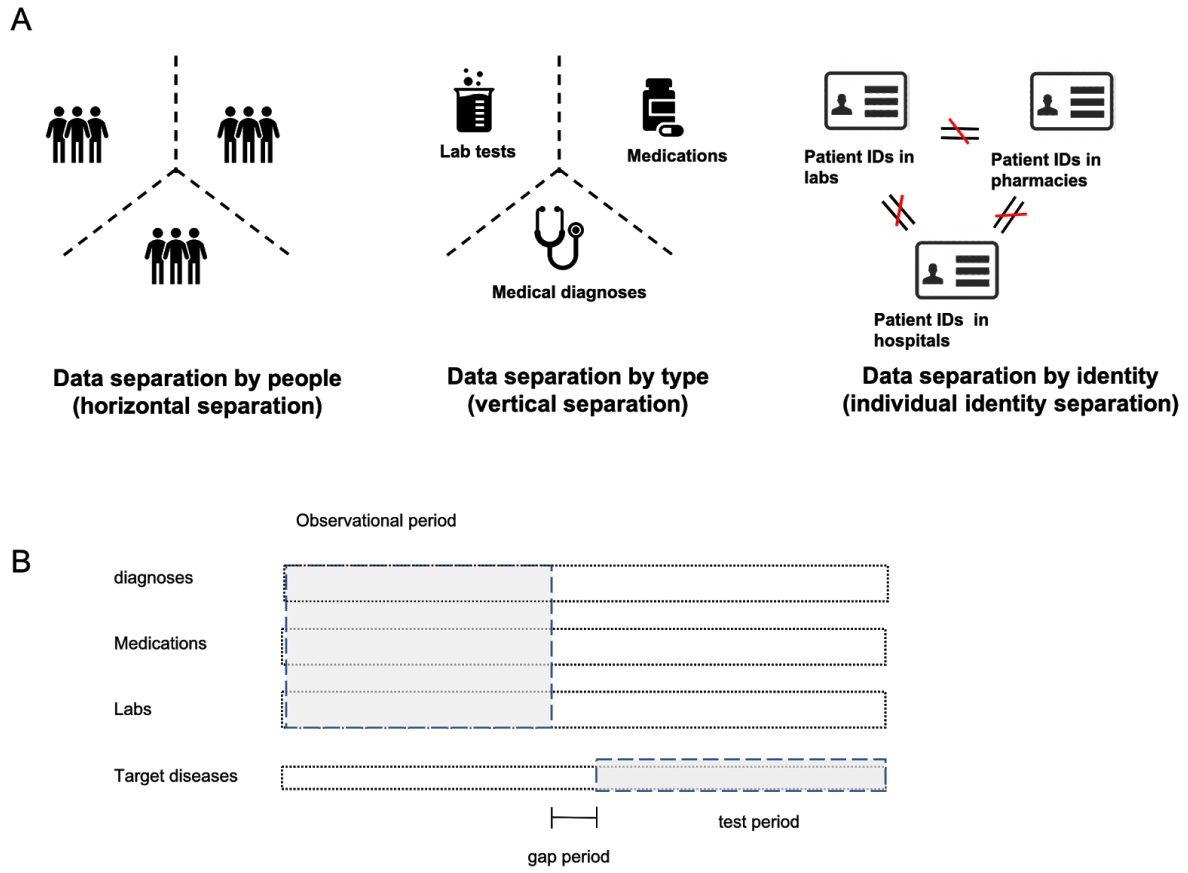


Figure 1. (A)Three degrees of separation. *Separation by individual* or *horizontal separation* refers to fragmentation of an individual's data across silos, for example across hospitals and clinics. *Separation by data type* or *vertical separation* refers to differences in the domain, semantics and structure of the data, for example, data from pharmacies, clinics and labs, each in their own nodes. *Separation by identity* refers to the problem of not being able to match individuals using their IDs, mostly due to inconsistency IDs among silos. **(B) Study period.** Patient's data are divided into three periods. Observational period is 24 months, gap period is 1 week and follow-up period is 23 months and 3 weeks. Diagnoses, medications and lab tests data of each patient in the observational period were used as predictive features for target diseases, such as diabetes in the test period. The 1-week gap period is introduced to avoid complications of encounters happening directly before diagnosis of target diseases.

Results

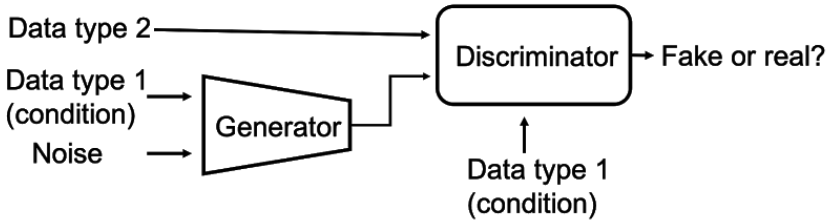
In order to train machine learning models on medical data separated by individual, by data type and by identity, A method, we referred to as confederated learning, was developed. Confederated learning consists of three steps: Step 1) Conditional generative adversarial networks were trained using data from the central analyzer to infer one data type from another, eg. inferring medications using diagnoses. Step 2) Missing data types from each silo were inferred using the model trained in step 1. Step 3) Task-specific models , such as a model to predict diagnoses of diabetes, were trained in a federated manner across all silos simultaneously. We conducted experiments to train disease prediction models using confederated learning on a large nationwide health insurance dataset from the U.S that is split into 99 silos. The models stratify individuals by their risk of diabetes, psychological disorders or ischemic heart disease in the next two years, using diagnoses, medication claims data and clinical lab test records of patients (See Methods section for details). The goal of these experiments is to test whether a confederated learning approach can *simultaneously* address all the three degrees of separation mentioned above.

Confederated learning. In this study, we assume there exists a central analyzer with a small amount connected data with all three data types and no separation at all. a large number data silos with only a single data type and all the three degrees of separation mentioned above (Figure 2). In *step 1*, using available data from the central analyzer,including diagnoses (X_{si}^{diag}),medications (X_{si}^{med}) and lab tests (X_{si}^{lab}), a cGAN was trained for each pair of data types. For example, inferring medications using diagnoses. To train the generator, two losses were used. A least square adversarial loss was included to minimize errors on discriminator. A L1 matching loss between the generated data and real data was included encourages the outputs to be close to observed data(Isola et al. 2017; Mao et al. 2019). A random Gaussian noise vector with length 100 was used as the source of stochasticity. In addition, another classifier was trained on the central analyzer to map each data type to the target disease label $Y_{si}^{disease}$, which will be used in step 2. There are two reasons why we used a cGAN based method instead of a deterministic supervised method to infer missing data type. Firstly,some of the individuals do not have all data types available in the data set. cGAN based methods are able to utilize these incomplete data for discriminator training. Secondly, we are more interested in the potential distribution of a data type rather than a point estimate due to the heterogeneous nature of healthcare. In *Step 2*, the cGANs and target disease classifiers obtained from step 1 were passed to each of the 99 silos. In each silo, X_{si}^{diag} ,X_{si}^{med} and X_{si}^{lab} ,if not available in the corresponding silo,were inferred using cGAN based on data type available (See Methods for details). The binary labels $Y_{si}^{disease}$ of targeted diseases were inferred from X using the classifier model obtained from step 1 .In *step 3*, after the data inference step, each silo has all three types of data , namely, diagnoses, medication and lab tests, as well as the inferred binary labels of targeted diseases. Therefore, a traditional federated learning method can be applied. We applied federated averaging

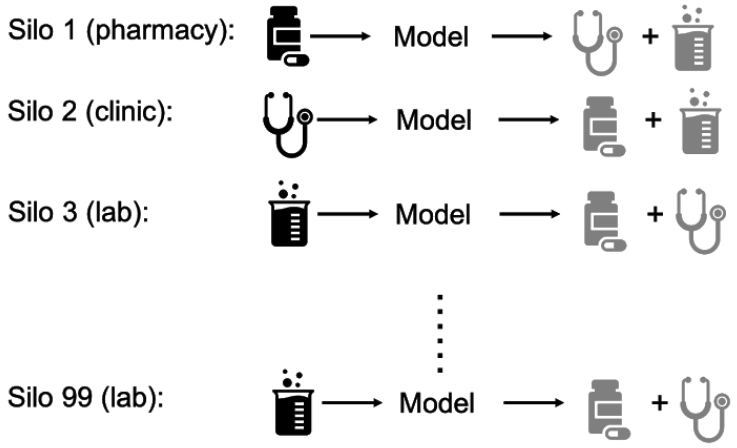
algorithms for step 3 due to its popularity and robustness (H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, Blaise Agüera y Arcas 2016).

Central analyzer has a small number of patients with all 3 data types and matched patient IDs

Step 1: Train conditional GAN to use one data type to infer the other two, for example:



Step 2: Infer data in each of the 99 silos using information available



Step 3: After data inference , conduct federated training for task specific models in each silo and aggregate the obtained algorithm

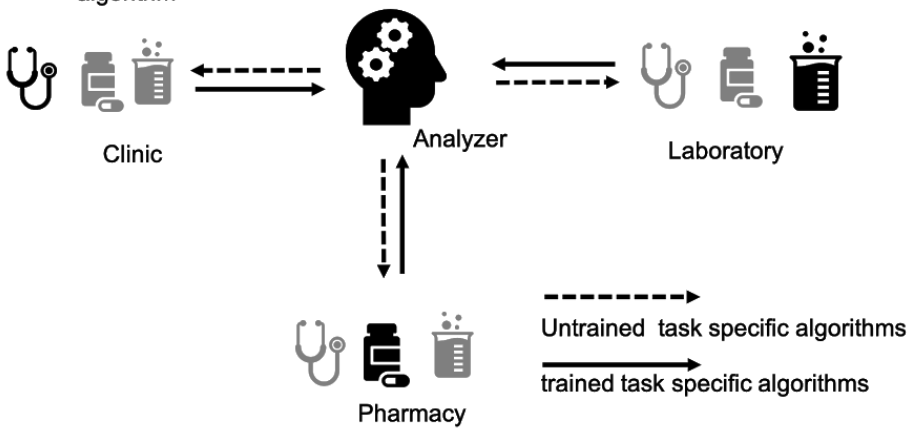


Figure 2. Confederated learning trains machine learning models using disconnected data separated by individual, data type and identity for Large-Scale Health System Intelligence. The method consists of 3 steps: Step 1) Conditional generative adversarial networks were trained using data from the central analyzer to infer one data type from another, eg. inferring medications using diagnoses. Step 2) Missing data types from each silo were inferred using the model trained in step 1. Step 3) Task-specific models , such as a model to predict diagnoses of diabetes, were trained in a federated manner across all silos simultaneously.

Disease-predicting models can be trained efficiently using confederated learning on disconnected data. We conducted predictive model training on health claim data in 99 silos separated by individual, by data type and by identity. Using confederated learning approach, the algorithm achieved an AUCROC of 0.79, AUCPR of 0.47, PPV of 0.56 and NPV of 0.81 predicting onset of diabetes in the test period, AUCROC of 0.72, AUCPR of 0.24, PPV of 0.36 and NPV of 0.91 predicting psychological disorders and AUCROC of 0.72, AUCPR of 0.24, PPV of 0.36 and NPV of 0.91 in predicting psychological disorders (Table 1).

In order to understand whether the confederated learning method efficiently utilized data separated by all three levels, we conducted 3 control experiments. Firstly ,when models were only trained on data from the central analyzer, the performance is inferior to models trained in a confederated manner (Table 1). Secondly, we conducted federated learning on across all silos with a specific data type (eg. only diagnoses) , the models obtained also show poorer results in all three tasks compared with confederated learning. Lastly, when we aggregated all the data from all silos such that data are not disconnected at all, the models trained in a centralized manner, as expected, show superior performances compared with those obtained from the confederated learning.

Models trained by confederated learning have good performance in individual silos. To better understand efficiency of confederated learning, performance of the disease predicting models in each silo was analyzed. The medical claims data in the silos come from different states in the U.S. (see Methods section for details). Therefore, the performance was measured at each of the 34 states included in this study. When predicting diabetes, AUCROC of model trained in confederated learning manner outperformed both model trained with only central analyzer's data and model trained by federated learning with a single data type in 19 out of the 34 states with 63.3% of individuals in the whole cohort (Figure 3 A). When predicting psychological disorders, confederated learning outperformed other two methods in 29 out of 34 states with 77.6% of individuals by AUCROC (Figure 3 B).When predicting ischemic heart disease, confederated learning outperformed other two methods in 20 out of 34 states with 56.6% of individuals by AUCROC (Figure 3 C). Similar patterns were observed when using AUCPR, PPV or NPV as performance metrics.

Table 1. Performance of disease predicting models trained using confederated learning on medical data distributed in 99 silos separated by individual, data type and identity.

	AUCROC	AUCPR	PPV	NPV
Data with no separation (centralized)				
Diabetes	0.82	0.52	0.65	0.81
Psychological disorders	0.75	0.26	0.35	0.91
Ischemic heart disease	0.72	0.18	0.20	0.91
Use only data with central analyzer				
Diabetes	0.77	0.39	0.41	0.81
Psychological disorders	0.63	0.15	0.21	0.91
Ischemic heart disease	0.68	0.16	0.18	0.91
Federated learning using one data type*				
Diabetes	0.77	0.45	0.54	0.81
Psychological disorders	0.58	0.12	0.13	0.90
Ischemic heart disease	0.66	0.15	0.18	0.91
Confederated learning				
Diabetes	0.79	0.47	0.56	0.81
Psychological disorders	0.72	0.24	0.36	0.91
Ischemic heart disease	0.70	0.17	0.20	0.91

*As models obtained from federated learning on diagnosis had better performance than models trained on medications or lab tests, only results of models trained on diagnoses are shown in this table.

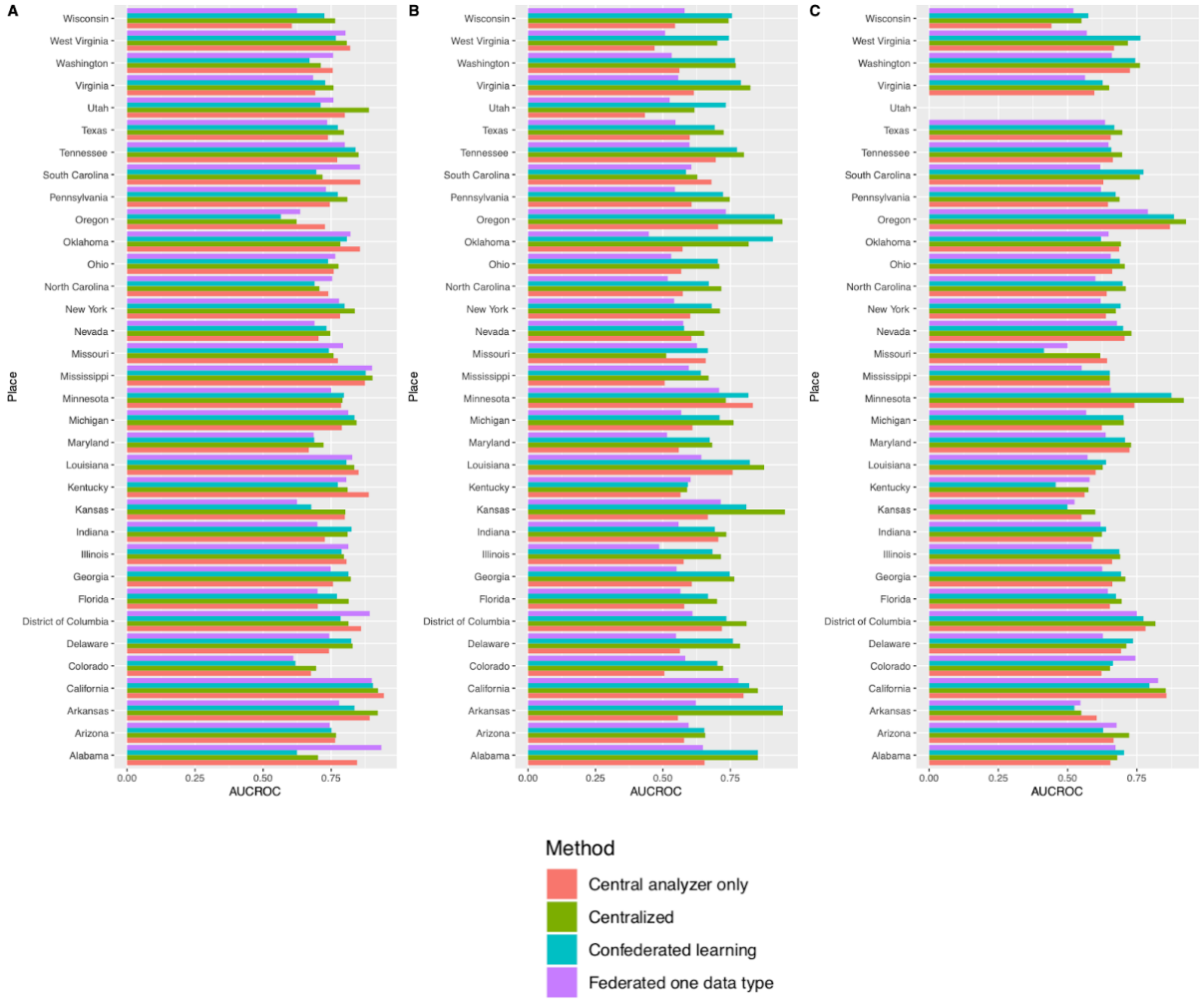


Figure 3. Performance of each method , by area under the receiver operating characteristic curve (AUCROC), in each state in the U.S. **(A)** Performance of model predicting diabetes **(B)** Performance of model predicting psychological disorders **(C)** Performance of model predicting ischemic heart disease.

Methods

Data source and cohort. The study uses claims data from a major U.S. health plan. Elements include the insurance plan type and coverage periods, age, sex, medications, and diagnoses associated with billed medical services, from July 1 2016 to June 31 2019. The dataset contains an indicator for insurance coverage by month. Only beneficiaries having full medical and pharmacy insurance coverage during the 48-month period were included. The study period is divided into a 24-month observational period, a 1 week gap period and a test period of 23 months and 3 weeks (Figure 1B). Individuals above age of 65 not enrolled in the Medicare Advantage program were excluded to ensure completeness of the private and public insurance data. In addition, members with pre-existing diagnoses of targeted disease within observational or gap were excluded. A total of 82,143 individuals were included in this study.

The input features to the confederated machine learning model include diagnoses as ICD 9 or ICD 10 codes, medications represented as National Drug Codes (NDC) and lab tests (encoded as LOINC codes). Lab test results were not available for this study. On average, each individual has 13.6 diagnoses, 6.9 prescriptions, and 7.4 LOINC codes during the 24 month observational period.

Study outcome. Diagnoses in the claim data were originally provided as an online International Classification of Diseases, Ninth Revision (ICD-9 and ICD-10). ICD9/10 codes of target diseases were selected according to mapping between ICD 9/10 codes and Phecodes from PheWASCatalog (https://phewascatalog.org/). Phecodes were defined by hierarchical grouping of ICD codes and were originally used for phenome-wide associations studies(Bastarache et al. 2018). For each member, we marked with a binary outcome variable (0 or 1) of whether a person had any claims related to each of the target diseases during the follow-up period. ICD-9/10 codes corresponding to PheCode 249-250 were used to define diabetes. ICD-9/10 codes corresponding to PheCode 295-306 were used to define psychological disorders and ICD-9/10 codes corresponding to PheCode 410-414.99 were used to define psychological disorders. Among the individuals, 16,824 had diabetes, 8,265 had psychological disorders and 8,044 had ischemic heart disease in the test periods.

Study setting. Data from 34 states in the U.S. were included into this study. The central analyzer has access to all three data types and ability to match patients' IDs across types from one state.In this study, data from California with information of 5433 individuals was used as the central analyzer's dataset .Data from the rest 33 states were divided into 99 silos by state and data type. We assume individual ID matching among silos was not possible, which is common is healthcare. For a specific silos $s \in \{1, 2, ...S\}$ with $S = 99$ in this study, Each individual i has either a diagnosis vector X_{si}^{diag} or a medication claim vector X_{si}^{med} from pharmacy, or lab test vector X_{si}^{lab} from clinical lab , where in each silo $i \in \{1, 2, ..., n^s\}$ with n^s being the number of beneficiaries in the silo. Our confederated learning method aims to train disease classification models using disconnected data from all the silos.

Claims for diagnoses, medications and lab tests during the observation period are the input features. The output of the classifier is a binary variable indicating whether the beneficiary had a target disease during the test period. Diabetes, psychological disorders and ischemic heart disease were chosen as targeted diseases in the test period due to their clinical importance. We simulated horizontal separation by separating the data for beneficiaries by U.S. state of residence. We simulated vertical separation by assuming that beneficiaries' diagnoses are only available in clinics, medication claims data are only kept in pharmacies and lab data only in labs. We simulated separation by identity by assuming individual ID matching was not possible among silos. Data is presumed to not be shared among different organizations nor across state lines. In total, besides the central analyzer, we simulated data

distributed across 99 distinct silos including 33 clinical silos, 33 pharmacy silos and 33 lab silos.

Model, training and performance evaluation details. Multi-layer neural network models with batch normalization and drop out were used for both generators and discriminators in the cGANs. Leaky ReLU was used as an activation function for hidden layers. A separate cGAN was built for each pair of data types. Another multi-layer neural network was used as classifiers to map each data type to binary labels. One classifier was built for each of the three disease labels.

In step 3, the goal of the training was to minimize the binary cross entropy, a metric for binary classification error, without moving any data out of their data silos. The objective function to minimize for classification model $f(X^{diag}, X^{med}, X^{lab}, \Theta)$ is:

$$L(X^{diag}, X^{med}, X^{lab}, \Theta) = \sum_{s=1}^{S} \sum_{i=1}^{n_s} -(Y_{si}\, log(f(X^{diag}, X^{med}, X^{lab}, \Theta)) + (1 - Y_{si}) log(1 - f(X^{diag}, X^{med}, X^{lab}, \Theta)))$$

Where Θ is the parameter of model f.

As data were not allowed to be moved out from their silos, it is not possible to train f by minimizing $L(X^{diag}, X^{med}, X^{lab}, \Theta)$ in a centralized manner. Therefore, we randomly initialized the parameters Θ as Θ_0 and sent model f and parameters Θ_0 to pharmacies or clinical labs in each silo $s \in S$. In each of the 99 silos, two of X^{diag}, X^{med} and X^{lab} are inferred from the third data type.

In the pharmacy silos, the loss function is then calculated as:

$$L(\widehat{X}^{diag}, X^{med}, \widehat{X}^{lab}, \Theta_{st}) = \sum_{i=1}^{n_s} -(Y_{si}\, log(f(\widehat{X}^{diag}, X^{med}, \widehat{X}^{lab}, \Theta_{st})) + (1 - Y_{si}) log(1 - f(\widehat{X}^{diag}, X^{med}, \widehat{X}^{lab}, \Theta_{st})))$$

Using stochastic gradient descent to minimize the loss, new parameters Θ_{st} were obtained. $t \in \{1, 2....T\}$ stands for number of global loops (Algorithm 1). n_s is the number of

individuals in the corresponding silo. Hats in $\widehat{X}^{diag}$ and $\widehat{X}^{lab}$ indicates that the diagnoses data and lab test data were inferred from medication data.

Similarly, In the lab silos:

$$L(\widehat{X}^{diag}, \widehat{X}^{med}, X^{lab}, \Theta_{st}) = \sum_{i=1}^{n_s} -(Y_{si} log(f(\widehat{X}^{diag}, \widehat{X}^{med}, X^{lab}, \Theta_{st})) + (1 - Y_{si}) log(1 - f(\widehat{X}^{diag}, \widehat{X}^{med}, X^{lab}, \Theta_{st})))$$

In the clinic silos:

$$L(X^{diag}, \widehat{X}^{med}, \widehat{X}^{lab}, \Theta_{st}) = \sum_{i=1}^{n_s} -(Y_{si} log(f(X^{diag}, \widehat{X}^{med}, \widehat{X}^{lab}, \Theta_{st})) + (1 - Y_{si}) log(1 - f(X^{diag}, \widehat{X}^{med}, \widehat{X}^{lab}, \Theta_{st})))$$

After Θ_{st} were trained locally in each single silos, they were sent back to the analyzer for aggregation by weighted averaging: $\Theta_t = \frac{1}{S} \sum_{s=1}^{S} \frac{n_s}{N}$ where N is the total number of beneficiaries included in the study from all states. Θ_t is then sent back to each silo to repeat the whole global cycle to obtain Θ_{t+1}. The global training cycles were stopped when the loss of the predictive model on the validation set did not decrease for 3 consecutive cycles.

20% of randomly chosen beneficiaries from all silos were reserved as test data, and not included in the training set, 20% individuals from central analyzer were chosen as validation set to adjust hyperparameters and the rest were used as training set. When conducting federated or confederated learning, data of 20% of beneficiaries from each node were used as an internal validation set. After hyperparameter tuning, both the training set and validation set were used to train the model to test performance.

Performance evaluation included area under the receiver operating characteristic curve (AUCROC) and area under the precision recall curve (AUCPR), AUCPR was used because the data are imbalanced--there are many more people without targeted diseases than those with the diseases. Instead of following the common practice of choosing a threshold that sets the false positive rates to be equal to the false negative rate (equal error rate), we chose the threshold which is 5% quantile of the predicted score of true positive in the validation set. We sought to favor a screening strategy and are willing to tolerate some false positives. Using this threshold, the positive predictive value (PPV) and negative predictive value (NPV) , which are commonly used metrics in clinical settings, were calculated (Table1) and used as performance metrics in addition to AUCROC and AUCPR.

Discussion

In this study, we demonstrated that health data distributed across silos separated by individual, data type and identity can be used to train machine learning models without

moving or aggregating data. Our method only showed a slight decrease in predictive accuracy in predicting risks of diabetes, psychological disorders or ischemic heart disease using previous diagnoses, medications and lab tests as inputs. We compared the performance of confederated learning approach with models trained on centralized data, only data with the central analyzer or a single data type across silos. The experimental results suggested that confederated learning trained predictive models efficiently across disconnected silos.

To the best of our knowledge, confederated learning proposed in this study, is the very first method developed to train machine learning models on data with the three degrees of separation mentioned above in healthcare setting. Compared with other methods for model training on horizontally and vertically separated data, this confederated learning algorithm does not require sophisticated computational infrastructure , such homomorphic encryption, nor frequent gradient exchange. In addition, due to the nature of the algorithm, the confederated training works fine even if a whole data type ,such as medication,is missing or not available for some patients.More importantly, patient matching or patient ID sharing is not required in our confederated setting, which are both difficult in many medical settings.

There are several limitations in this study. Firstly , our method allows model training without patient ID matching at the price of requiring the central analyzer to have some matched data and ignoring some interactions among different data types. In our disease prediction example, the central analyzer had fully connected data from one of the 34 states and the separation did not cause significant performance reduction likely because the cGAN models trained were able to perform relatively inference and each single data type had already caught a decent amount of information internally. However, in some other applications, more direct inter-data type interaction might be crucial. Secondly, we demonstrate performance using a single data source due to limited medical data availability. To make a more generalizable conclusion, more data sets will be needed in the experiments.

We anticipate that this confederated approach can be extended to more degrees of separation. Other types of separation, such as separation by temporality , separation by insurance plan, separation by healthcare provider can all be potentially be explored using confederated learning strategy.

References

al., Bonawitz Keith et. 2016. “Practical Secure Aggregation for Federated Learning on User-Held Data.” *arXiv*.

al., Konečný Jakub et. 2015. “Federated Optimization: Distributed Optimization Beyond the Datacenter.” *arXiv*.

Cano, Ignacio, Markus Weimer, Dhruv Mahajan, Carlo Curino, and Giovanni Matteo Fumarola. 2016. “Towards Geo-Distributed Machine Learning.” *arXiv Preprint*.

Diederik P. Kingma, Jimmy Ba. 2014. “Adam: A Method for Stochastic Optimization.”

arXiv.
Dollard, Joanne, Christopher Barton, Jonathan Newbury, and Deborah Turnbull. 2012. "Falls in Old Age: A Threat to Identity." *Journal of Clinical Nursing*. https://doi.org/10.1111/j.1365-2702.2011.03990.x.
"Fatal Injury Data | WISQARS | Injury Center | CDC." 2019. January 18, 2019. https://www.cdc.gov/injury/wisqars/fatal.html.
"Federated Learning: Collaborative Machine Learning without Centralized Training Data." n.d. Google AI Blog. Accessed March 25, 2019. http://ai.googleblog.com/2017/04/federated-learning-collaborative.html.
Friedman, Charles P., Adam K. Wong, and David Blumenthal. 2010. "Achieving a Nationwide Learning Health System." *Science Translational Medicine* 2 (57): 57cm29.
Geyer, R. C., Klein, T., & Nabi, M. 2017. "Differentially Private Federated Learning: A Client Level Perspective." *arXiv:*
Haasum, Ylva, and Kristina Johnell. 2017. "Use of Antiepileptic Drugs and Risk of Falls in Old Age: A Systematic Review." *Epilepsy Research*. https://doi.org/10.1016/j.eplepsyres.2017.10.022.
H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, Blaise Agüera y Arcas. 2016. "Communication-Efficient Learning of Deep Networks from Decentralized Data." *arXiv*.
Heinrich, S., K. Rapp, U. Rissmann, C. Becker, and H-H König. 2010. "Cost of Falls in Old Age: A Systematic Review." *Osteoporosis International: A Journal Established as Result of Cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA* 21 (6): 891–902.
Homer, Mark L., Nathan P. Palmer, Kathe P. Fox, Joanne Armstrong, and Kenneth D. Mandl. 2017. "Predicting Falls in People Aged 65 Years and Older from Insurance Claims." *The American Journal of Medicine* 130 (6): 744.e17–744.e23.
Hopewell, Sally, Olubusola Adedire, Bethan J. Copsey, Graham J. Boniface, Catherine Sherrington, Lindy Clemson, Jacqueline Ct Close, and Sarah E. Lamb. 2018. "Multifactorial and Multiple Component Interventions for Preventing Falls in Older People Living in the Community." *Cochrane Database of Systematic Reviews* 7 (July): CD012221.
Hu, Guoqing, and Susan P. Baker. 2012. "An Explanation for the Recent Increase in the Fall Death Rate among Older Americans: A Subgroup Analysis." *Public Health Reports* 127 (3): 275–81.
Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. "Image-to-Image Translation with Conditional Adversarial Networks." *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2017.632.
Kosse, N. M., K. Brands, J. M. Bauer, T. Hortobagyi, and C. J. C. Lamoth. 2013. "Sensor Technologies Aiming at Fall Prevention in Institutionalized Old Adults: A Synthesis of Current Knowledge." *International Journal of Medical Informatics* 82 (9): 743–52.
Kramarow, Ellen, Li-Hui Chen, Holly Hedegaard, and Margaret Warner. 2015. "Deaths from Unintentional Injury among Adults Aged 65 and over: United States, 2000-2013." *NCHS Data Brief*, no. 199 (May): 199.
Lord, Stephen R., Catherine Sherrington, Hylton B. Menz, and Jacqueline C. T. Close. 2007. *Falls in Older People: Risk Factors and Strategies for Prevention*. Cambridge University Press.
Mandl, Kenneth D., Isaac S. Kohane, Douglas McFadden, Griffin M. Weber, Marc Natter,

Joshua Mandel, Sebastian Schneeweiss, et al. 2014. “Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS): Architecture.” *Journal of the American Medical Informatics Association: JAMIA* 21 (4): 615–20.

Mao, Xudong, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. 2019. “On the Effectiveness of Least Squares Generative Adversarial Networks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/tpami.2018.2872043.

McMurdo, Marion E. T., Angela M. Millar, and Fergus Daly. 2000. “A Randomized Controlled Trial of Fall Prevention Strategies in Old Peoples’ Homes.” *Gerontology*. https://doi.org/10.1159/000022139.

Moyer, Virginia A., and U.S. Preventive Services Task Force. 2012. “Prevention of Falls in Community-Dwelling Older Adults: U.S. Preventive Services Task Force Recommendation Statement.” *Annals of Internal Medicine* 157 (3): 197–204.

Overstall, P. W. 1985. “Epidemiology and Pathophysiology of Falls.” *Fits, Faints and Falls in Old Age*. https://doi.org/10.1007/978-94-009-4892-1_3.

Panel on Prevention of Falls in Older Persons, American Geriatrics Society and British Geriatrics Society. 2011. “Summary of the Updated American Geriatrics Society/British Geriatrics Society Clinical Practice Guideline for Prevention of Falls in Older Persons.” *Journal of the American Geriatrics Society* 59 (1): 148–57.

Praneeth, {vepakomma, Otkrist Gupta, Tristan Swedish, and Ramesh Raskar}. 2018. “Split Learning for Health: Distributed Deep Learning without Sharing Raw Patient Data.” *arXiv*.

Qi, Jinwei, Xin Huang, and Yuxin Peng. 2017. “Cross-Media Retrieval by Multimodal Representation Fusion with Deep Networks.” *Communications in Computer and Information Science*. https://doi.org/10.1007/978-981-10-4211-9_22.

Stephen, {hardy, Wilko Henecka, Hamish Ivey-Law, Richard Nock, Giorgio Patrini, Guillaume Smith, and Brian Thorne}. 2017. “Private Federated Learning on Vertically Partitioned Data via Entity Resolution and Additively Homomorphic Encryption.” *arXiv*.

Verma, Santosh K., Joanna L. Willetts, Helen L. Corns, Helen R. Marucci-Wellman, David A. Lombardi, and Theodore K. Courtney. 2016. “Falls and Fall-Related Injuries among Community-Dwelling Adults in the United States.” *PloS One* 11 (3): e0150939.

Yang, Yu, Xinhua Hu, Qiang Zhang, and Rui Zou. 2016. “Diabetes Mellitus and Risk of Falls in Older Adults: A Systematic Review and Meta-Analysis.” *Age and Ageing*. https://doi.org/10.1093/ageing/afw140.

Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, Vikas Chandra. 2018. “Federated Learning with Non-IID Data.” *arXiv*, June.

Zhai, Xiaohua, Yuxin Peng, and Jianguo Xiao. 2014. “Learning Cross-Media Joint Representation With Sparse and Semisupervised Regularization.” *IEEE Transactions on Circuits and Systems for Video Technology*. https://doi.org/10.1109/tcsvt.2013.2276704.

Zhang, Haijun, and Nanfeng Xiao. 2015. “Learning Hierarchical Representations for Face Recognition Using Deep Belief Network Embedded with Softmax Regress and Multiple Neural Networks.” *Proceedings of the 2015 2nd International Workshop on Materials Engineering and Computer Sciences*. https://doi.org/10.2991/iwmecs-15.2015.1.