SCALING NEUROSCIENCE RESEARCH USING FEDERATED LEARNING

Dimitris Stripelis, José Luis Ambite

Information Sciences Institute
University of Southern California
Marina del Rey, CA, USA

Pradeep Lam, Paul Thompson

Imaging Genetics Center
USC Stevens Neuroimaging and Informatics Institute
Marina del Rey, CA, USA

ABSTRACT

The amount of biomedical data continues to grow rapidly. However, the ability to analyze these data is limited due to privacy and regulatory concerns. Machine learning approaches that require data to be copied to a single location are hampered by the challenges of data sharing. Federated Learning is a promising approach to learn a joint model over data silos. This architecture does not share any subject data across sites, only aggregated parameters, often in encrypted environments, thus satisfying privacy and regulatory requirements. Here, we describe our Federated Learning architecture and training policies. We demonstrate our approach on a brain age prediction model on structural MRI scans distributed across multiple sites with diverse amounts of data and subject (age) distributions. In these heterogeneous environments, our Semi-Synchronous protocol provides faster convergence.

Index Terms— Deep Learning, Federated Learning

1. INTRODUCTION

Advances in computer technology, electronic medical records, and cost-efficient biomedical data acquisition, from sensors to genetic tests, allow healthcare organizations and biomedical research studies to collect increasing amounts of data. Analysis of these vast datasets using machine learning approaches promises novel discoveries. Unfortunately, privacy, security, and regulatory constraints make sharing datasets across studies or organizations extremely difficult, so that this promise is largely unfulfilled since joint analyses are limited.

To address these challenges, Federated Learning [1, 2] has emerged as a novel privacy-preserving distributed machine learning paradigm that enables large-scale cross-institutional analysis without the need to move the data out of its original location. Federated Learning allows institutions to collaboratively train a machine learning model (e.g., a neural network) by aggregating the parameters (e.g., gradients) of local models trained on local data. Since subject data is not shared, and parameters can be protected through encryption, privacy concerns are ameliorated. Even though Federated Learning was originally developed for mobile and edge devices, it is being increasingly applied in biomedical and healthcare domains [3, 4, 5, 6, 7]. We developed a Federated Learning architecture and training policies resilient to data and computational heterogeneity, where different sites may have different data amounts, target distributions, and computational capabilities [8, 9], which are often characteristic of biomedical studies.

Brain age prediction from brain structural MRIs is a challenging biomedical task. The difference between the predicted and chronological brain age values is a phenotype related to aging and brain disease. Recent work [10, 11] has shown that deep learning methods can accurately predict an individual's brain age. However, data scarcity limits the power of these methods, since privacy requirements make data sharing difficult. The Federated Learning paradigm is a natural fit for these challenging learning environments.

Here, we present our Federated Learning (FL) architecture and an empirical evaluation of brain age prediction under homogeneous, and heterogeneous environments with different amounts of data, and with data not independently and identically distributed (Non-IID) across sites [2, 8, 9, 12]. We compare the effectiveness of the federated model to its centralized counterpart. We show that in heterogeneous environments, our communication-efficient Semi-Synchronous training policy [8] provides faster convergence.

2. RELATED WORK

Federated Learning holds much promise in healthcare domains [6]. Silva et al. [7] present an open-source FL framework for healthcare, supporting different models and optimization methods. FL has been used for phenotype discovery [13], for patient representation learning [14], and for identifying similar patients across institutions [3].

In biomedical imaging, Federated Learning has been applied to multiple tasks, including whole-brain segmentation of MRI T1 scans [15], brain tumor segmentation [4, 16], multi-site fMRI classification and identification of disease biomarkers [17], and for identification of brain structural relationships across diseases and clinical cohorts using (federated) dimensionality reduction from shape features [5]. COINSTAC [18] provides a privacy-preserving distributed data processing framework for brain imaging.

Depending on the requirements and computational characteristics of the federated learning environment, the participating sites (learners) can be organized under different topologies [2, 6]. In a star topology (Figure 1), the learners communicate through a head server that is responsible for coordinating the federation training (e.g., our work and [4, 16, 17]). In a peer-to-peer topology [15], learners can communicate directly with each other without a distinguished coordinator.

Most current FL approaches [1, 12, 22] compute the global model through a synchronous communication protocol. However, in heterogeneous environments, stragglers may slow down convergence. Our Semi-Synchronous protocol [8] counters this inefficiency by assigning more local computation to the underutilized learners to accelerate convergence.

Preserving data privacy is critical in Federated Learning environments. Common methods for privacy protection are differential privacy [19] and homomorphic encryption [20]. For example, the federated learning system for brain tumor segmentation in [16] used differential privacy techniques. We are developing a homomorphic encryption approach in our architecture, but it is out of the scope of this paper.

Finally, our deep learning model for estimating brain age from structural MRI scans is closely related (albeit different) to [10, 11, 21]. However, we use Federated Learning to learn the joint model, as opposed to using majority voting and linear regression data blending to combine CNN brain age predictions from different data sources as in [11].

3. FEDERATED LEARNING

Federated Learning operates over sites that do not share data, so the joint model is obtained through parameter sharing and mixing [2, 6]. Figure 1 shows the typical Federated Learning architecture. Here, we consider that a single neural network, known to all the sites, is being optimized. Each site (learner) trains the neural network on its own local private dataset and shares only the locally-learned parameters with a head server, the Federation controller, which is responsible for aggregating the local parameters to generate a community neural network, which is in turn sent back to the learners.

Federated Optimization. The primary goal in Federated Learning is to jointly learn a global/community model across a federation of N learners, with non-co-located data, by optimizing the global objective $f(w)$:

$$w^* = \underset{w}{\arg\min} f(w) \; \text{ with } \; f(w) = \sum_{k=1}^{N} \frac{p_k}{\sum p_k} F_k(w) \quad (1)$$

where p_k is the contribution value of a learner k to the federation and $F_k(w)$ its local objective function. The contribution value p_k can be static, or dynamically defined at run time [9]. In much recent work [1, 22], the learners are weighted based on the number of local training examples ($p_k = |D_k^T|$), since

this is a good proxy for the value of a local model, but other methods that directly measure performance are possible [9].
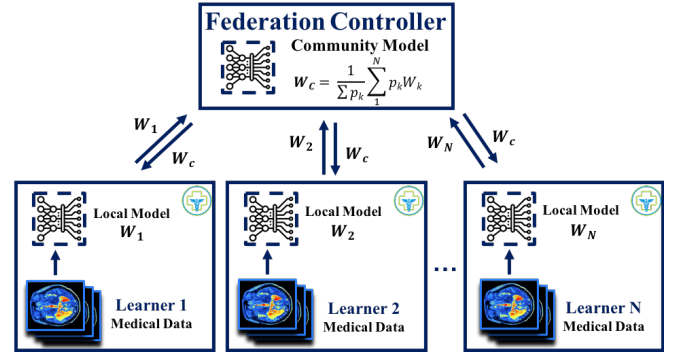

Fig. 1: Federated Learning Architecture

Synchronous Training. In the original Federated Learning algorithm [1, 22], every learner performs a predefined number of local updates (batches or epochs) before reaching a synchronization point where it shares its local model with the federation. This computational approach, which we refer to as Synchronous Federated Average (*SyncFedAvg*), has been extensively explored [1, 12, 22].

Semi-Synchronous Training. We introduced a Semi-Synchronous training protocol (*SemiSync*) [8] where each learner trains for a given amount of time before synchronization. Each learner processes a variable number of data batches between synchronization points depending on its computational power and amount of data. SemiSync parameterizes the synchronization period based on the time that it takes for the slowest learner in the federation to perform a single epoch. The number of local updates (batches) $\mathcal{B}_k$ a learner k performs between synchronization points is computed as:

$$t_{max}(\lambda) = \lambda * \max_{k \in N} \left\{ \frac{|D_k^T|}{\beta_k} * t_{\beta_k} \right\}, \lambda, \beta_k, t_{\beta_k} > 0$$
$$\mathcal{B}_k = \frac{t_{max}}{t_{\beta_k}}, \quad \forall k \in N \quad (2)$$

where $|D_k^T|$ is the number of local training examples, β_k is the learner's local batch size defined at global model initialization and t_{β_k} is the time it takes to perform a local batch (i.e., an update). The hyperparameter λ controls the communication frequency of the learners by adjusting the number of local updates per learner based on the time it takes for the slowest learner to perform λ local epochs.

This training policy is particularly effective in federated learning settings where learners have homogeneous computational power, but heterogeneous amounts of data, as well as in settings where learners have heterogeneous computational power and/or heterogeneous amounts of data [8].

4. FEDERATED BRAIN AGE PREDICTION

The learning task we investigate here is brain age prediction. Deep 3D convolutional regression networks have been used for brain age prediction [10, 11]. These networks extend the VGG and ResNet architectures to 3D images by replacing 2D convolution/maxpool operations with their 3D counterparts.

Neural Architecture. Figure 2 shows the convolutional encoding network we trained for the brain age prediction task. The model architecture is similar to that in [21] with the main difference being the replacement of the batch normalization (BatchNorm) layer with an instance normalization (InstanceNorm) layer. Collectively, the network consists of seven blocks, with the first five composed of a 3x3x3 3D convolutional layer (stride=1, padding=1), followed by an instance norm, a 2x2x2 max-pool (stride=2), with ReLU activation functions. The number of filters in the first block is 32 (and doubles until 256) with both layers 4 and 5 having 256 filters. The sixth block contains a 1x1x1 3D convolutional layer (stride=1, filters=64), followed by an instance norm and ReLU activation. The final, seventh, block contains an average pooling layer, a dropout layer (set to p=0.5 during training), and a 1x1x1 3D convolutional layer (stride=1). To train the model we used Mean Squared Error as loss function and Vanilla SGD as the network's optimizer.
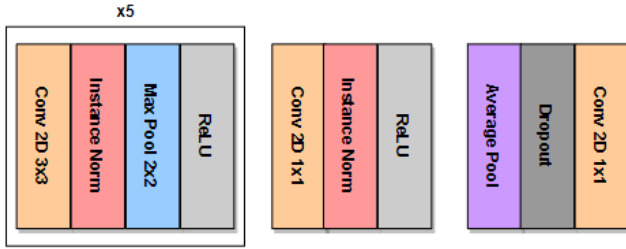

Fig. 2: BrainAgeCNN

Federated Model. During federated training all learners train on their local data using the same neural architecture and hyperparameters (e.g., learning rate, batch size). Once a learner finishes its local training, it sends its local model parameters to the controller.

5. EXPERIMENTS

Our goal is to apply Federated Learning to hospital consortia, and to large research studies like Enigma (enigma.ini.usc.edu). As an initial step in a controlled environment, we analyzed brain MRI data from the UK Biobank [23], a large epidemiological study of 500,000 people residing in the United Kingdom, some with neuroimaging. We explored several heterogeneous federated learning scenarios with different data distributions and amounts of data per learner and evaluated the performance and convergence rate of the federation.

NeuroImaging Data. From the original UKBB dataset of 16,356 individuals with neuroimaging, we selected 10,446 who had no indication of neurological pathology, and no psychiatric diagnosis as defined by the ICD-10 criteria. The age range was 45-81 years (mean: 62.64; SD: 7.41; 47% women, 53% men). All image scans were evaluated with a manual quality control procedure, where scans with severe artifacts were discarded. The remaining scans were processed using a standard preprocessing pipeline with non-parametric intensity normalization for bias field correction1 and brain extraction using FreeSurfer and linear registration to a (2 mm)^3 UKBB minimum deformation template using FSL FLIRT. The final dimension of the registered images was 91x109x91. The 10,446 records were split into 8356 for train and 2090 for test.
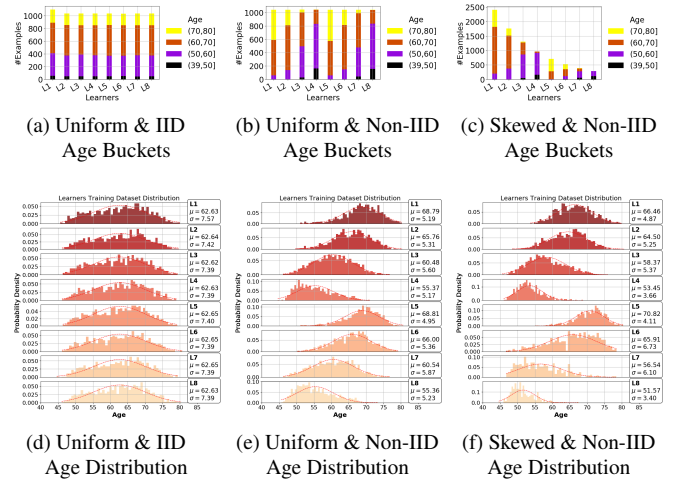

(a) Uniform & IID Age Buckets

(b) Uniform & Non-IID Age Buckets

(c) Skewed & Non-IID Age Buckets

(d) Uniform & IID Age Distribution

(e) Uniform & Non-IID Age Distribution

(f) Skewed & Non-IID Age Distribution

Fig. 3: UKBB Federation Data Distributions

Data Distributions. We define several challenging learning environments by partitioning the centralized UKBB neuroimaging training dataset (8356 records) across a federation of 8 learners.[1] Every environment was evaluated on the same test dataset (2090 records) and the learners used their allocated records for training. As shown in Figure 3: panels (a, b, c) show the amount of data (and age buckets), and the corresponding (d, e, f) show the detailed age distribution of each of the 8 learners. Figures 3(a,d) show a uniform (same amount of data per learner) and IID (all ages) distribution. Figures 3(b,e) show a uniform, but non-IID (subset of ages) distribution. Figures 3(c,f) show a skewed (different amount of data per learner) and non-IID distribution.

Training Environment. We established a federation of 8 learners by assigning one learner to each GPU of a server with 8 GeForce GTX 1080 Ti graphics cards (10 GB RAM each), 40 Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz, and 128GB DDR4 RAM. All learners trained on the same CNN model (Fig. 2). For both centralized and federated models, we

---

[1]Available at: `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/2RKAQP`

used Vanilla SGD with a learning rate of $5\mathrm{x}10^{-5}$ and a batch size (β_k) of 1. For SyncFedAvg, each learner runs 4 local epochs in all distributions. For SemiSyncFedAvg, the time per batch (t_{β_k}) for every learner is 120ms, the maximum time of a single epoch across all learners is 280secs (learner 1 that holds the largest partition: $\sim$2,400 examples), thus for $\lambda = 4$ the maximum time (t_{max}) is 1,120secs (cf. Eq. 2). Finally, the assigned number of local updates ($\mathcal{B}_k$) per learner is close to 9300. For all experiments, the random seed was 1990.
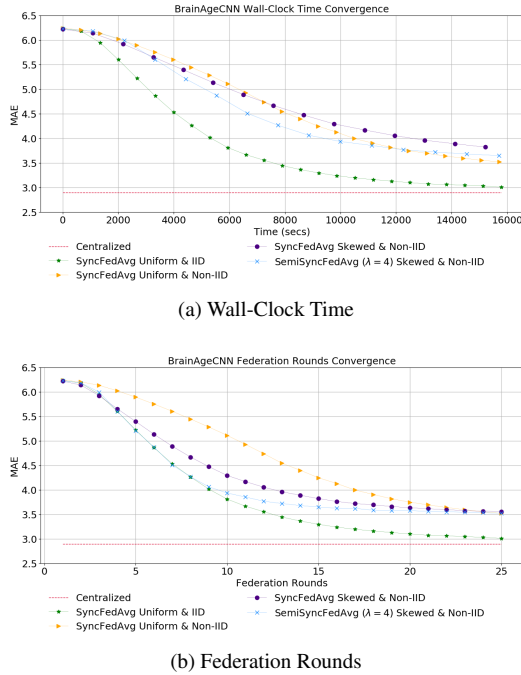

(a) Wall-Clock Time

(b) Federation Rounds

Fig. 4: UKBB Brain Age Federated Policies Performance

Evaluation. We compare the performance of brain age prediction in settings with uniform and skewed data sizes, and with IID and non-IID age distributions. Since we consider a homogeneous computational environment with 8 identical learners (GPUs) and the computational cost of each learner only depends on the amount of data, for uniform data sizes we only show results for the Synchronous Federated Average policy (SemiSync performs the same). For skewed data sizes, we show both synchronous and semi-synchronous policies.

Figure 4a shows the performance of the training policies over the different environments in terms of elapsed (wall-clock) time (i.e., the 8 learners running in parallel). For the Uniform and IID setting, the federation reaches a Mean Absolute Error (MAE) value that is very close to the one achieved by the centralized model. More challenging Non-IID settings lower the performance, with a final error of 0.5 years over the centralized model (cf. Table 1). The Semi-Synchronous FedAvg policy has a faster convergence and slightly better final performance than Synchronous FedAvg. Even though the computational power of each learner is the same (identical

GPUs), in the skewed data setting for the SyncFedAvg policy the learners with the smaller amounts of data remain idle until the learner with the most data finishes its allocated epochs. In contrast, the SemiSync policy continuously processes batches without any idle time. This additional computation results in the improved convergence in this setting.

Figure 4b shows the performance in terms of federation rounds, which is a proxy for the communication cost of the policy.[2] At the end of the allocated epochs for SyncFedAvg, or λ time for SemiSyncFedAvg, all learners share their models with the Federation Controller, which then computes the community model and sends it back to the learners. Thus, the number of models exchanged through the network is twice the number of federation rounds times the number of learners. As before, SemiSyncFedAvg shows faster convergence in terms of communication cost than SyncFedAvg.

		MSE	RMSE	MAE	Corr
Centralized Model		12.885 ± 0.021	3.589 ± 0.003	2.895 ± 0.006	0.881
Federated Model Data Distribution	**Policy**				
Uniform & IID	SyncFedAvg	13.749 ± 0.138	3.707 ± 0.018	2.995 ± 0.018	0.875
Uniform & Non-IID	SyncFedAvg	19.853 ± 1.347	4.453 ± 0.151	3.625 ± 0.135	0.861
Skewed & Non-IID	SyncFedAvg	19.148 ± 0.086	4.376 ± 0.009	3.553 ± 0.003	0.851
	SemiSync(λ=4)	18.491 ± 0.122	4.311 ± 0.015	3.505 ± 0.008	0.864

Table 1: UKBB Evaluation. Mean and std values for 3 runs.

6. CONCLUSION

We have demonstrated the effectiveness of the Federated Learning paradigm in the neuroimaging domain by collaboratively learning a global model for brain age prediction. We empirically evaluated the convergence of the federated model in statistically heterogeneous learning environments. Our immediate future work includes investigating additional learning tasks, such as disease prediction, and incorporating homomorphic encryption in the architecture We simulated the non-IID case by subsampling UKB data, but we plan to examine the more realistic case where learners receive data from different scanning protocols and cohorts, which would exhibit natural acquisition and population differences. This will better show the different relative performance of the different training policies. Finally, we plan to explore federated transfer learning in neuroimaging.

7. ACKNOWLEDGMENTS

---

[2]Note that the wall-clock time for each federation round depends on the data distribution. To process the 25 federation rounds in Figure 4b, SyncFedAvg takes 15,482 seconds in the Uniform and IID setting, 15,777 seconds in the Uniform and Non-IID setting, and 22,713 seconds in the Skewed and Non-IID setting. For 25 federations rounds in the Skewed and Non-IID setting SemiSyncFedAvg takes 23,048 seconds.

8. COMPLIANCE WITH ETHICAL STANDARDS

This is a study of previously collected, anonymized, de-identified data, available in a public repository. Data access approved by UK Biobank under Application Number 11559.

9. REFERENCES

[1] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *AISTATS*. PMLR, 2017.

[2] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.

[3] Junghye Lee, Jimeng Sun, Fei Wang, Shuang Wang, et al., "Privacy-preserving patient similarity learning in a federated environment: development and analysis," *JMIR Medical Informatics*, vol. 6, no. 2, pp. e20, 2018.

[4] Micah J Sheller, et al., "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in *Intl. MICCAI Brainlesion Workshop*. Springer, 2018.

[5] Santiago Silva, Boris A Gutman, Eduardo Romero, Paul M Thompson, Andre Altmann, and Marco Lorenzi, "Federated learning in distributed medical databases: Meta-analysis of large-scale subcortical brain data," in *16th Intl. Symp. on Biomedical Imaging*, IEEE 2019.

[6] Nicola Rieke, Jonny Hancox, Wenqi Li, Fausto Milletari, et al., "The future of digital health with federated learning," *npj Digital Medicine*, vol. 3, no. 119, 2020.

[7] Santiago Silva, Andre Altmann, et al., "Fed-biomed: A general open-source frontend framework for federated learning in healthcare," in *Intl. MICCAI DART/DCL Workshop*. Springer, 2018

[8] Dimitris Stripelis and José Luis Ambite, "Semi-synchronous federated learning," *arXiv:2102.02849*, 2021.

[9] Dimitris Stripelis and José Luis Ambite, "Accelerating federated learning in heterogeneous data and computational environments," *arXiv:2008.11281*, 2020.

[10] James H Cole, Rudra PK Poudel, Dimosthenis Tsagkrasoulis, et al., "Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker," *NeuroImage*, vol. 163, pp. 115–124, 2017.

[11] Benedikt Atli Jónsson, Gyda Bjornsdottir, TE Thorgeirsson, et al., "Brain age prediction using deep learning uncovers associated sequence variants," *Nature Communications*, vol. 10, no. 1, pp. 1–10, 2019.

[12] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith, "Federated optimization in heterogeneous networks," *MLSys*, 2020.

[13] Dianbo Liu, Dmitriy Dligach, and Timothy Miller, "Two-stage federated phenotyping and patient representation learning," in *BioNLP Workshop*, 2019.

[14] Yejin Kim, Jimeng Sun, Hwanjo Yu, and Xiaoqian Jiang, "Federated tensor factorization for computational phenotyping," in *ACM SIGKDD*, 2017, pp. 887–895.

[15] Abhijit Guha Roy, Shayan Siddiqui, et al., "Braintorrent: A peer-to-peer environment for decentralized federated learning," *arXiv:1905.06731*, 2019.

[16] Wenqi Li, et al., "Privacy-preserving federated brain tumour segmentation," in *Intl. Workshop on Machine Learning in Medical Imaging*. Springer, 2019.

[17] Xiaoxiao Li, Yufeng Gu, Nicha Dvornek, et al., "Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results," *Medical Image Analysis*, vol. 65, 101765, 2020.

[18] Sergey M. Plis, et al., "COINSTAC: A privacy enabled model and prototype for leveraging and processing decentralized brain imaging data," *Frontiers in Neuroscience*, vol. 10, pp. 365, 2016.

[19] Martin Abadi, Andy Chu, et al., "Deep learning with differential privacy," in *ACM SIGSAC Conference on Computer and Communications Security*, 2016.

[20] Chengliang Zhang, et al., "Batchcrypt: Efficient homomorphic encryption for cross-silo federated learning," in *USENIX Annual Technical Conference*, 2020.

[21] Han Peng, Weikang Gong, Christian F. Beckmann, Andrea Vedaldi, and Stephen M. Smith, "Accurate brain age prediction with lightweight deep neural networks," *Medical Image Analysis*, p. 101871, 2020.

[22] Keith Bonawitz, Hubert Eichner, et al., "Towards federated learning at scale: System design," *MLSys*, 2019.

[23] Karla L Miller, Fidel Alfaro-Almagro, et al., "Multimodal population brain imaging in the UK biobank prospective epidemiological study," *Nature Neuroscience*, vol. 19, no. 11, pp. 1523–1536, 2016.