Brain Age Estimation Using Structural MRI: A Clustered Federated Learning Approach

Seyyed Saeid Cheshmi[1], Abtin Mahyar[1], Anita Soroush[1], Zahra Rezvani[2,3], and Bahar Farahani[4]
[1]Department of Computer and Data Sciences, Shahid Beheshti University, Tehran, Iran
[2]School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
[3]Department of Computer Science, Faculty of Mathematical Sciences, Alzahra University, Tehran, Iran
[4]Cyberspace Research Institute, Shahid Beheshti University, Tehran, Iran

***Abstract*—Estimating brain age based on structural Magnetic Resonance Imaging (MRI) is one of the most challenging and prominent areas of research in recent medical imaging and neuroscience studies. The significance of brain age prediction in the early diagnosis of neurological disorders has fueled a resurgence of interest in this field. Various studies have addressed this issue using a spectrum of techniques, from traditional machine learning to deep neural networks. The majority of these techniques employ centralized paradigms, which do not adequately preserve privacy. To tackle this problem, a handful of studies have utilized a federated approach. In this study, we propose a novel hierarchical clustered federated learning approach that carefully captures and considers the similarities of the clients' predictions on a certain benchmark dataset. This method enhances performance in Non-Independent and Identically Distributed (Non-IID) environments while preserving privacy. We use a multi-site dataset that provides a broad variety of MRI scans, characterizing a proper Non-IID environment. Our method achieves a Mean Absolute Error (MAE) of 3.86, while the non-clustered FedAvg federated approach attains a 4.14 MAE on the test set.** [1]

***Index Terms*—Brain Age Estimation, Privacy-Preserving Machine Learning (PPML), Federated Learning (FL), Clustering.**

I. Introduction

Predicting brain age from structural Magnetic Resonance Imaging (MRI) scans and, subsequently, using it for Brain-Age-Gap Estimation (BrainAGE) has turned out to be of great importance in age-related disorders. For an individual, the gap between their predicted brain age and chronological brain age is considered a valuable biomarker for prognosticating the risk of some brain disorders, including, but not limited to, Alzheimer's disease, schizophrenia, and bipolar disorder [1].

Deep Learning (DL) and Machine Learning (ML) have been widely applied to different medical tasks, demonstrating competitive performance compared to humans. Recent works [2] have shown that DL methods can predict brain age accurately. However, DL models are data-hungry; for more precise prediction, they need a vast amount of data. Unfortunately, gathering extensive medical data may be challenging due to privacy concerns, as well as legal and ethical restrictions on data sharing. Regulatory frameworks, such as Health Insurance Portability and Accountability Act (HIPAA) or General Data Protection Rule (GDPR), restrict sharing of personally identifiable health information without the individual's authorization. To tackle this issue, Privacy-Preserving Machine Learning (PPML) techniques, including Federated Learning (FL) [3] have been developed. FL allows for extensive cross-institutional analysis without requiring data to be transferred from its original location to the center. FL facilitates the training of a model by aggregating the parameters of local models that are trained on local data. Since the aggregation is done just by sharing parameters, not the actual data, privacy concerns are relieved.

There is a growing trend to apply FL in the areas of biomedical research and healthcare [4]. In recent years, few FL studies have been proposed for brain age estimation. Their main focus was encryption and privacy preservation [5], [6], while federated aggregation methods and configurations were not discussed considerably. To address the shortcomings of existing techniques, we propose a cluster-based FL approach, which has reasonable performance in Non-Independent and Identically Distributed (Non-IID) environments. We conduct various experiments for the brain age prediction task on Open Big Healthy Brains (OpenBHB) [7] dataset. We use three types of data distributions for evaluation: centralized, decentralized in an IID environment, and decentralized in a Non-IID environment. We aim to use the FL approach to predict brain age and improve the results when the data distribution is heterogeneous. Thus, the main contribution of this work is to introduce a clustered FL approach based on the similarities of the clients' predictions on a certain benchmark dataset for a regression task, brain age estimation.

The rest of the paper is organized as follows: Section II explains related work, section III introduces the proposed method, section IV discusses our experiments and results, and Section V presents the conclusion.

II. Related Work

The issue of calculating brain age has been addressed in several studies in the literature using various approaches. They can be categorized into traditional and cutting-edge research. Earlier studies have used ML techniques, namely, linear regression, SVR, GPR, and RVR, to try and solve this task [8]. To reduce the size of the input photos, reduce computing costs, and improve model performance, they have experimented with

979-8-3503-4647-3/23/$31.00 ©2023 IEEE

[1]Code is available at https://github.com/Abtinmy/Clustered-FL-BrainAGE.

a variety of strategies, such as various preprocessing steps, feature selection, and extraction. Contrarily, recent research that made use of convolution neural networks and their widely accepted designs attempted to address this issue [2]. We primarily concentrate on DL models since they often surpass traditional ones in terms of performance and have a greater capacity for task generalization.

A summary of related works for brain age estimation is presented in Table I. These studies can be split into two subgroups: centralized and decentralized. In the centralized paradigm, the whole data is stored in a single site. Based on the input data, these methods themselves can be split into two categories. In the first category, the age of the brain is inferred using only a 2D slice of the entire structural MRI scans. The correlation between the various slices in the entire brain image is not taken into account in these works. Despite this, they perform rather well when compared to 3D models, which need more expensive computations and parameters. [9] tries to utilize the correlation between the slices using LSTM modules. They begin by extracting encoding sequences using a 2D Convolutional Neural Network (CNN) encoder and then attempt to estimate the age of the brain by finding relationships between extracted encodings. In [10], [11], they input 2D slices of the source images to the VGG, and AlexNet architectures, respectively, and then use the derived features to conduct the estimation. The relationship between several patches in the 2D-sliced input picture and the entire image will be computed in [12] using multiple attention modules and then utilized to estimate age by lowering the loss that comprises multiple predictions for each relation.

In the second category, the models use the whole 3D scan (voxel-based) as input. [13], which is an extended version of [14], proposed a 3D CNN model that perceives the whole brain image as the input and segments it into patches and then utilizes Pearson's correlation to determine how similar they are to one another. In [15], [16], 3D VGG-based models are used by replacing 2D convolutional layers with 3D ones. The most significant contributions made by these works included the region-wise study of the brain [15] and a unique model inspired by VGGNet (SFCN [16]). Similar contributions were performed by using ResNet-based 3D models in [17], [18]. Several research attempted to improve their performance in a variety of ways, such as performing an ensemble model of the 3D CNNs architecture [19], [20], developing innovative models based on DenseNet [21], [22], and utilizing auto-encoders [23]. In a similar study to [12], the relationship between image pairs was identified using Transformer, and four relations terms were anticipated by the model [24].

In the decentralized setting, data is distributed across multiple sites, and studies only concern voxel-based data. In [5], a framework for secure FL using fully-homomorphic encryption was produced. They evaluated their model on the brain age prediction task. In [25], another FL architecture was presented and was applied in [6], where MetisFL was introduced and evaluated on the brain age prediction task. In these works, [5] and [6], the main focus was to tackle privacy-preserving issues and deal with membership and model inversion attacks with the use of homomorphic encryption.

Although there are some other enhanced studies on cluster-based federated learning in other fields, the main focus of our work is dedicated to brain age estimation. In [26], they presented a new iterative-based approach, Iterative Federated Clustering Algorithm (IFCA), It alternates between estimating cluster identities and minimizing loss functions. The technique has two variants: gradient averaging, in which the gradients of local loss functions are computed and shared, and model averaging, in which local devices perform stochastic gradient descent and share updated models with the central machine for averaging. Cluster assignments are improved iteratively by the method until convergence. Also, [27] introduced another approach based on the inference similarity of the clients using an adjacency matrix in classification tasks. Our work, in particular, proposes an approach for cluster-based federated learning in a regression task on a large-scale brain age dataset.

III. Proposed Method

We utilize an FL approach to tackle the problem of brain age prediction, preserving privacy, and boosting the model's performance in a Non-IID environment. In this regard, we assume a secure connection between clients and the server is established to alleviate the concerns of vulnerabilities with respect to privacy. One of the challenges in performing the FL process is the aggregation method which highly affects the performance of the procedure. Recent studies demonstrated that having a single global model does not necessarily enhance performance [33], [34]. One of the alternative approaches is to perform clustering on the clients and group them; then, weight aggregation is performed in each cluster. Most studies conduct this clustering by the weights that are transferred to the server from the clients [35].

Here we proposed our FL approach, which clusters the clients with a brand-new method. In this method, each local model and the global model for each cluster have the same 3D-based architecture and it is further explained in the following four phases (See Fig. 1):

1) Initially, a subset of the training dataset is sampled which consists of multiple records from all age ranges to represent the wide variety of scans and be suitable for benchmarking. This set is achieved by binning the training dataset into uniform bins, randomly choosing a few records from each bin, and excluding them from the training set. After that, this subset is stored in the server. This dataset is referred to *Benchmark Dataset* for simplification. It is worth mentioning since this data is consistent and only compromises a scant proportion of the dataset, so privacy would not be at risk. Also, synthesized data can be used as an alternative.
2) Then, local models train without any aggregation for a few epochs. This is done so that local models have the opportunity to learn significant patterns in their data and adapt to its data distribution.
3) Subsequently, clients send their parameters back to the server, then the server uses the benchmark dataset to

TABLE I: Related studies on brain age estimation.

Paradigm	Data Type	Ref.	Modality	Model[a]	#Samples	MAE[b]	Year	Dataset
Centralized	Slice-based	[10]	T1	VGG	1099	4	2017	ABIC
		[14]	T1	CNN	484	4.7	2019	ABIDE, ADNI, BNU, ICBM, IXI
		[9]	-	RNN	10446	2.86	2020	UK Biobank
		[11]	T1	AlexNet	594	4.51	2021	ADNI GO, ADNI 2, IXI, OASIS
		[12]	T1	Attention + Transformers	8379	2.7	2021	BGSP, OASIS-3, NIH-PD, ABDIE-I, IXI, DLBS, CMI, CoRR
	Voxel-based	[28]	T1	CNN	5496	4.45	2019	Rotterdam Study
		[29]	T1	CNN	562	5	2020	IXI
		[15]	T1	VGG	10158	4.06	2020	Cam-CAN, IXI, SALD, DLBS, OASIS-1, CoRR, SchizConnect, ADNI, AIBL, OASIS-2, PPMI, NIFD, BGSP, SLIM
		[19]	T1	Ensemble of DL	10176	3.07	2020	ADNI, PPMI, ICBM, AIBL, SLIM, OASIS, CANDI, IDA, COBRE, CNP, CORR, FCP
		[17]	T1	ResNet	21382	2.87	2020	UK Biobank
		[18]	T1	ResNet	12864	2.84	2021	German National Cohort
		[13]	T1	CNN	1016	2.19	2021	ABIDE
		[16]	T1	VGG	14503	2.14	2021	UK Biobank
		[22]	T1	DenseNet	4127	4.2	2022	Mayo
		[23]	T1	Autoencoder	2074	3.85	2022	CNNT1, CNNTOF
		[20]	T1	Ensemble of DL + ML	2641	3.19	2022	PAC2019, BANC, STORM
		[21]	T2	DenseNet	23302	2.97	2022	KCH, GSTT
		[30]	T1	CNN	3609	2.85	2022	IXI, SALD, NKI, CoRR, UKB, PNC, 973, HCP
		[24]	T1	Attention + Transformers	6049	2.38	2022	TMGHBCH, NIH-PD, ABIDE-I, BGSP, BeijingEN, IXI, DLBS, OASIS
		[31]	T1	ConvNeXt	11728	2.09	2022	ABDIE I, ABDIE II, ADNI, CoRR, DLBS, ICBM, IXI, NKI-RS, OASIS-3, OpenfMRI, SALD
Decentralized	Voxel-based	[25]	T1	CNN (Federated)	10446	2.99	2021	UK Biobank
		[32]	T1	SVR	-	2.58	2022	UPENN-PNC, UK Biobank
		[6]	T1	CNN (Federated)	3312	2.9	2022	ADNI 1, ADNI 2, ADNI 3, OASIS, AIBL

[a]Backbones, variations, or modules the study used is reported
[b]Lowest MAE is reported based on the paradigm of the study

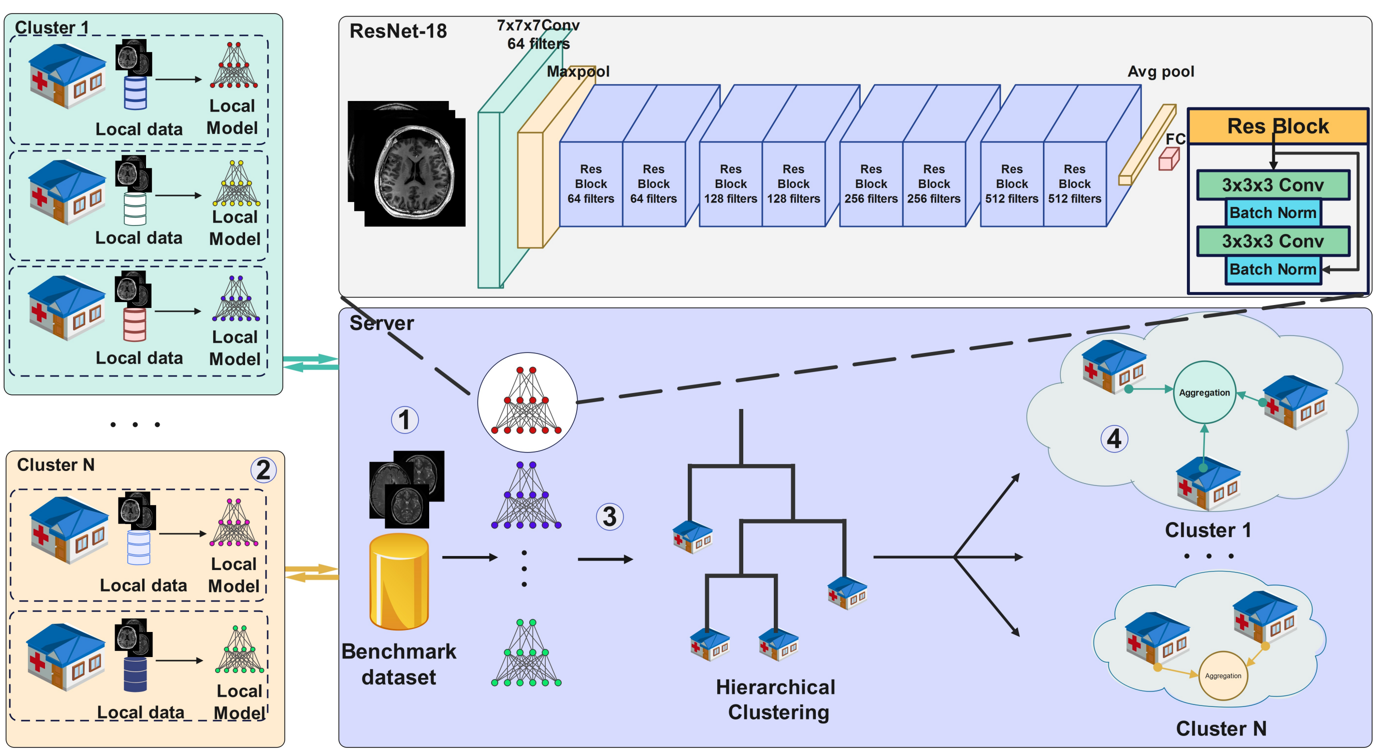


Fig. 1: The proposed architecture of cluster-based FL approach for brain age estimation.

infer and get predictions from the clients. Clients' predictions are represented as a vector in the inference time and the server clusters the clients based on the distances of these vectors. In other words, clients that have similar outputs, group together. The Agglomerative clustering algorithm is used in this phase, which is a bottom-up hierarchical approach to group the clients. Starting with every data point in its own cluster, it iteratively combines the two nearest clusters using a selected linkage criterion. The linkage is set to *Ward*, which aims to decrease the variance of the merged clusters. This approach typically results in compact, spherical clusters of relatively similar size. This clustering procedure is repeated every ω epoch. The algorithm is executed for different numbers of clusters to find the suitable one.

4) Afterwards, the parameter aggregation will perform separately on models presented in each cluster using the FedAvg algorithm in the server. In FedAvg, the weighted average of models' parameters in a cluster determines the global model parameters. Then, local models will sync with the global model. The federated aggregation will be repeated for more iterations within each cluster until the convergence.

In circumstances where the data distribution of the clients is not steady and changes over time, such as online learning, the clusters may alter sometimes. To deal with this issue, phases 3 and 4 are performed repeatedly.

For a better understanding of the procedures being done in the server and clients, the steps are presented in Algorithm 1 and Algorithm 2, respectively. K represents the number of clusters, $\{W_i\}_{i=1}^m$ are the weights belong to the i^{th} client, and $\{\tilde{W}^{(k)}\}_{k=1}^K$ are the weights of the global model of the k^{th} cluster. Algorithm 1 is responsible for conducting clustering and aggregating the weights of each cluster, while Algorithm 2 trains each local model with its corresponding data for a certain number of steps π.

IV. Experiments

A. Dataset

We conduct our experiments on the OpenBHB dataset [7]. In this large-scale dataset, 5330 3D brain MRI scans are gathered from 71 different acquisition sites, from which 3984 scans are publicly available (3227 for the train set, 757 for the validation set, which consists of 362 internal tests, and 390 external tests). OpenBHB is made up of 10 datasets. The subjects included in OpenBHB come from European-American, European, and Asian genetic backgrounds, which provides more diversity. In this dataset, age ranges from 5 to 88 years old, and sex distribution is well balanced for all ages Fig. 2a. Three preprocessing pipelines have been applied to all these datasets. As a result, three modalities are available, derived from the same T1-w MRI scans: VBM (Voxel-Based Morphometry maps), SBM (Suface-Based Morphometry indices), and quasi-raw (simple linear alignment of images). Quasi-raw is the modality that we are going to work on. In this modality, images went through some preprocessing steps, starting with ANTS Bias Field Correction, followed by FLS FLIRT with 9 degrees of freedom, affine registration to the $1mm^3$, and non-brain tissues removal using a brain mask respectively.

Algorithm 1 Server side procedure

Initialize:
  K, $\{W_i\}_{i=1}^m$, $\{\tilde{W}^{(k)}\}_{k=1}^K$, ω,
  $clustered \leftarrow False$, $BenchmarkDataset$
for $Epoch = 1, \ldots N$ **do**
  if $clustered$ **then**
    for each cluster $k = 1, \ldots K$ **do**
      $\tilde{W}^{(k)} \leftarrow FedAvg(W_i \in C_k)$
      for $W_i \in C_k$ **do**
        $W_i \leftarrow \tilde{W}^{(k)}$
        $Update\_Client(W_i)$
      end for
    end for
  else
    for $w \in W_i$ **do**
      $Update\_Client(w)$
    end for
  end if
  if $Epoch \,\%\, \omega \;=\; 0$ **then**
    for each client $i = 1, \ldots m$ **do**
      $pred_i \leftarrow Inference(C_i, BenchmarkDataset)$
    end for
    $AgglomerativeClustering(K, pred)$
    $clustered \leftarrow True$
  end if
end for

Algorithm 2 Update_Client

Inputs:
  W_{client}
for $Step$ = 1 to π **do**
  Update W_{client} with local data.
end for

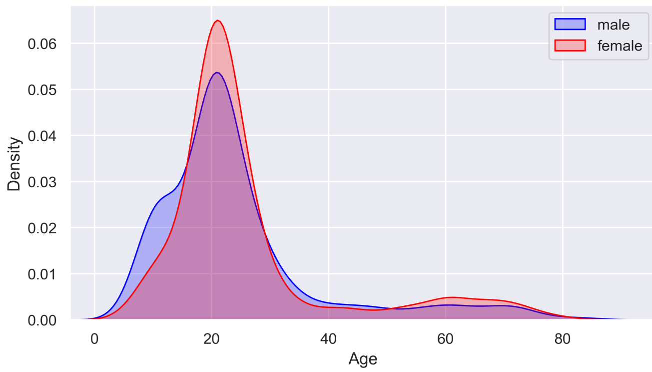
B. Data Distributions

We report our results in three paradigms.

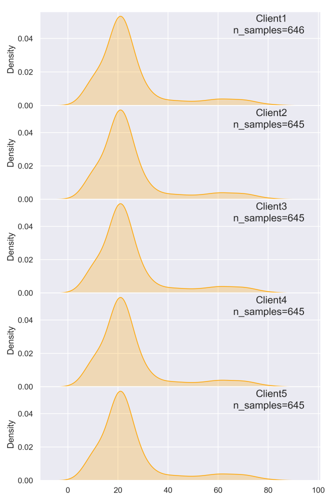
1) Centralized: The model is trained in a centralized paradigm on the OpenBHB dataset without division. The dataset is intact, and we do not use FL for training and evaluating the models. We conduct these experiments for comparison purposes.

2) Decentralized in an IID Environment: The dataset is partitioned across five clients. To simulate an IID environment, the dataset is split into five subsets so that they are all identical in terms of population and distribution. Fig. 2b illustrates the age distribution of each client.

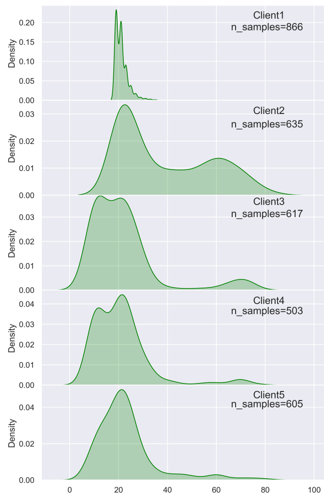
3) Decentralized in a non-IID Environment: In general, data heterogeneity, or non-IID data distribution, is referred to as a decentralized paradigm when the features of the clients' datasets have different distributions. When it comes to brain age estimation, these features may refer to geographical location, gender, acquisition site, etc. Particularly, in our work, we consider data heterogeneity as the difference in



(a) Centralized paradigm.



(b) Decentralized paradigm in an IID environment.



(c) Decentralized paradigm in a non-IID environment.

Fig. 2: Different data distribution in each paradigm.

age distribution between the dataset of each client. Again, we consider five clients, but this time, the dataset is split so that each pair of resulting distributions are different. Moreover, these datasets have different numbers of scans as well. This non-IID environment is an acceptable simulation of real-world situations. Fig. 2c demonstrates the age distribution of each client in this paradigm.

C. Training Criteria

As mentioned earlier, we use different data distributions for each paradigm. For model evaluation, we use the internal test set, including records belonging to sites used in training. We use Adam as the optimizer with an initial learning rate of 1e-4 and train each model for 45 epochs. We also choose a batch size of 4 for training both paradigms. In the centralized paradigm, we employ a learning rate scheduler that decays the learning rate every ten epochs by half. In the federated paradigm, we have 150 steps in each epoch, and the local batch size $B = \lfloor \frac{N_m}{I} \rfloor$ which N_m represents the number of scans in the m^{th} client and I is the number of iterations. In each epoch, local parameters aggregate according to the aggregation pace π, which is set to 50. Furthermore, we set ω to 5, performing clustering every five epochs.

D. Result

The evaluation of the proposed method and other baselines on train and test sets are reported in Table II. According to the

TABLE II: Brain age estimation results on train and test sets using different models and paradigms.

Paradigm	Model[a]	MAE		Clusters
		Train	Test	
Centralized	ResNet-18 [36]	0.54	2.86	-
	BotNet50 [37]	3.01	3.42	
	EfficientNet [38]	1.05	3.62	
Federated IID	ResNet-18	0.92	2.92	-
Federated non-IID	ResNet-18	1.35	4.14	-
IFCA [26]	ResNet-18	1.78	4.91	{1,2,3,4,5},{},{}
Federated non-IID Cluster-based (Ours)	ResNet-18 K = 2	1.05	3.87	{1,2},{3,4,5}
	ResNet-18 K = 3	1.06	3.99	{1},{2},{3,4,5}
	ResNet-18 K = 4	0.91	3.86	{1},{2},{3,4},{5}

[a] 2D convolution layers were replaced with 3D ones in these architectures

results of centralized models, we chose ResNet-18 as our base model for FL experiments. Results demonstrated that our introduced method outperforms the FL process with non-clustered FedAvg aggregation in the non-IID setting. Moreover, we execute our method with three different numbers of clusters. As can be concluded from Table II, this parameter affects the performance of the model and should be tuned according to the corresponding data. Furthermore, as shown in Table, our method performs better clustering than IFCA [26], such that final clusters have clients whose data distribution is similar. MAE on test set per federated learning rounds is illustrated for federated paradigm experiments in Fig. 3. According to this, in a non-IID environment, the cluster-based federated models have lower MAE compared to other approaches. The lowest MAE belongs to the cluster-based approach with $K = 4$.

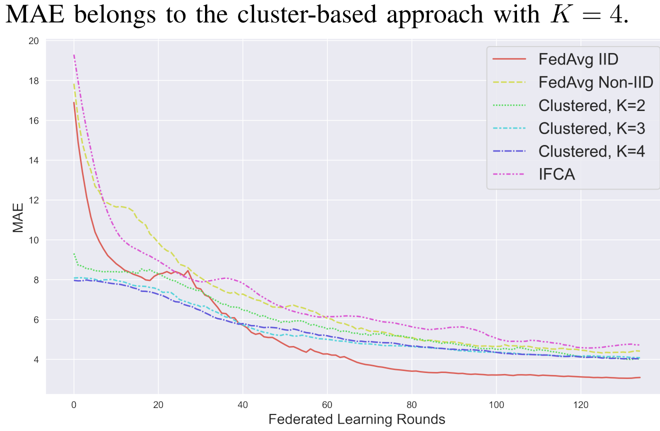


Fig. 3: Mean absolute error during federated learning rounds.

V. Conclusion

This paper suggests a novel cluster-based method for dealing with heterogeneous data, particularly in a healthcare-related task, brain age estimation. We have also demonstrated that our approach outperforms non-clustered FedAvg on non-IID settings and performs quite well compared to centralized and decentralized IID. Nevertheless, this clustering approach requires complete client participation during training, which is a challenging problem. Future efforts might also propose a creative solution to adequately accommodate new clients, which was not presented in the FL training process. Moreover, this technique demands that the number of clusters is explicitly selected. Thus a solution to this problem might be offered.

REFERENCES

[1] C. Gaser, K. Franke, S. Klöppel, N. Koutsouleris, H. Sauer, and ADNI, "Brainage in mild cognitive impaired patients: Predicting the conversion to alzheimer's disease," *PLoS ONE*, vol. 8, p. e67346, 06 2013.

[2] M. Tanveer, M. Ganaie, I. Beheshti, T. Goel, N. Ahmad, K.-T. Lai, K. Huang, Y.-D. Zhang, J. Del Ser, and C.-T. Lin, "Deep learning for brain age estimation: A systematic review," *arXiv preprint arXiv:2212.03868*, 2022.

[3] F. Firouzi, B. Farahani, M. Barzegari, and M. Daneshmand, "Ai-driven data monetization: The other face of data in iot-based smart and connected health," *IEEE Internet of Things Journal*, vol. 9, no. 8, pp. 5581–5599, 2020.

[4] F. Firouzi, S. Jiang, K. Chakrabarty, B. Farahani, M. Daneshmand, J. Song, and K. Mankodiya, "Fusion of iot, ai, edge–fog–cloud, and blockchain: Challenges, solutions, and a case study in healthcare and medicine," *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 3686–3705, 2022.

[5] D. Stripelis, H. Saleem, T. Ghai, N. Dhinagar, U. Gupta, C. Anastasiou, G. Ver Steeg, S. Ravi, M. Naveed, P. M. Thompson *et al.*, "Secure neuroimaging analysis using federated learning with homomorphic encryption," in *17th International Symposium on Medical Information Processing and Analysis*, vol. 12088. SPIE, 2021, pp. 351–359.

[6] D. Stripelis, U. Gupta, H. Saleem, N. Dhinagar, T. Ghai, R. Sanchez, C. Anastasiou, A. Asghar, G. V. Steeg, S. Ravi *et al.*, "Secure federated learning for neuroimaging," *arXiv preprint arXiv:2205.05249*, 2022.

[7] B. Dufumier, A. Grigis, J. Victor, C. Ambroise, V. Frouin, and E. Duchesnay, "Openbhb: a large-scale multi-site brain mri data-set for age prediction and debiasing," *NeuroImage*, vol. 263, p. 119637, 2022.

[8] S. Mishra, I. Beheshti, and P. Khanna, "A review of neuroimaging-driven brain age estimation for identification of brain disorders and health conditions," *IEEE Reviews in Biomedical Engineering*, 2021.

[9] P. K. Lam, V. Santhalingam, P. Suresh, R. Baboota, A. H. Zhu, S. I. Thomopoulos, N. Jahanshad, and P. M. Thompson, "Accurate brain age prediction using recurrent slice-based networks," in *16th International Symposium on Medical Information Processing and Analysis*, vol. 11583. SPIE, 2020, pp. 11–20.

[10] T.-W. Huang, H.-T. Chen, R. Fujimoto, K. Ito, K. Wu, K. Sato, Y. Taki, H. Fukuda, and T. Aoki, "Age estimation from brain mri images using deep learning," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 849–852.

[11] L. Lin, G. Zhang, J. Wang, M. Tian, and S. Wu, "Utilizing transfer learning of pre-trained alexnet and relevance vector machine for regression for predicting healthy older adult's brain age from structural mri," *Multimedia Tools and Applications*, vol. 80, pp. 24 719–24 735, 2021.

[12] S. He, P. E. Grant, and Y. Ou, "Global-local transformer for brain age estimation," *IEEE transactions on medical imaging*, vol. 41, no. 1, pp. 213–224, 2021.

[13] L. Bellantuono, L. Marzano, M. La Rocca, D. Duncan, A. Lombardi, T. Maggipinto, A. Monaco, S. Tangaro, N. Amoroso, and R. Bellotti, "Predicting brain age with complex networks: From adolescence to adulthood," *NeuroImage*, vol. 225, p. 117458, 2021.

[14] N. Amoroso, M. La Rocca, L. Bellantuono, D. Diacono, A. Fanizzi, E. Lella, A. Lombardi, T. Maggipinto, A. Monaco, S. Tangaro *et al.*, "Deep learning and multiplex networks for accurate modeling of brain age," *Frontiers in aging neuroscience*, vol. 11, p. 115, 2019.

[15] X. Feng, Z. C. Lipton, J. Yang, S. A. Small, F. A. Provenzano, A. D. N. Initiative, F. L. D. N. Initiative *et al.*, "Estimating brain age based on a uniform healthy population with deep learning and structural magnetic resonance imaging," *Neurobiology of aging*, vol. 91, pp. 15–25, 2020.

[16] H. Peng, W. Gong, C. F. Beckmann, A. Vedaldi, and S. M. Smith, "Accurate brain age prediction with lightweight deep neural networks," *Medical image analysis*, vol. 68, p. 101871, 2021.

[17] A. Kolbeinsson, S. Filippi, Y. Panagakis, P. M. Matthews, P. Elliott, A. Dehghan, and I. Tzoulaki, "Accelerated mri-predicted brain ageing and its associations with cardiometabolic and brain disorders," *Scientific Reports*, vol. 10, no. 1, pp. 1–9, 2020.

[18] L. Fisch, J. Ernsting, N. R. Winter, V. Holstein, R. Leenings, M. Beisemann, K. Sarink, D. Emden, N. Opel, R. Redlich *et al.*, "Predicting brain-age from raw t 1-weighted magnetic resonance imaging data using 3d convolutional neural networks," *arXiv preprint arXiv:2103.11695*, 2021.

[19] G. Levakov, G. Rosenthal, I. Shelef, T. R. Raviv, and G. Avidan, "From a deep learning model back to the brain—identifying regional predictors and their relation to aging," *Human brain mapping*, vol. 41, no. 12, pp. 3235–3252, 2020.

[20] Z. Zhang, R. Jiang, C. Zhang, B. Williams, Z. Jiang, C.-T. Li, P. Chazot, N. Pavese, A. Bouridane, and A. Beghdadi, "Robust brain age estimation based on smri via nonlinear age-adaptive ensemble learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2146–2156, 2022.

[21] D. A. Wood, S. Kafiabadi, A. Al Busaidi, E. Guilhem, A. Montvila, J. Lynch, M. Townend, S. Agarwal, A. Mazumder, G. J. Barker *et al.*, "Accurate brain-age models for routine clinical mri examinations," *Neuroimage*, vol. 249, p. 118871, 2022.

[22] J. Lee, B. J. Burkett, H.-K. Min, M. L. Senjem, E. S. Lundt, H. Botha, J. Graff-Radford, L. R. Barnard, J. L. Gunter, C. G. Schwarz *et al.*, "Deep learning-based brain age prediction in normal aging and dementia," *Nature Aging*, vol. 2, no. 5, pp. 412–424, 2022.

[23] P. Mouches, M. Wilms, D. Rajashekar, S. Langner, and N. D. Forkert, "Multimodal biological brain age prediction using magnetic resonance imaging and angiography with the identification of predictive regions," *Human Brain Mapping*, vol. 43, no. 8, pp. 2554–2566, 2022.

[24] S. He, Y. Feng, P. E. Grant, and Y. Ou, "Deep relation learning for regression and its application to brain age estimation," *IEEE transactions on medical imaging*, vol. 41, no. 9, pp. 2304–2317, 2022.

[25] D. Stripelis, J. L. Ambite, P. Lam, and P. Thompson, "Scaling neuroscience research using federated learning," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 1191–1195.

[26] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran, "An efficient framework for clustered federated learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 19 586–19 597, 2020.

[27] M. Morafah, S. Vahidian, W. Wang, and B. Lin, "Flis: Clustered federated learning via inference similarity for non-iid data distribution," *IEEE Open Journal of the Computer Society*, vol. 4, pp. 109–120, 2023.

[28] J. Wang, M. J. Knol, A. Tiulpin, F. Dubost, M. de Bruijne, M. W. Vernooij, H. H. Adams, M. A. Ikram, W. J. Niessen, and G. V. Roshchupkin, "Gray matter age prediction as a biomarker for risk of dementia," *Proceedings of the National Academy of Sciences*, vol. 116, no. 42, pp. 21 213–21 218, 2019.

[29] N. Pardakhti and H. Sajedi, "Brain age estimation based on 3d mri images using 3d convolutional neural network," *Multimedia Tools and Applications*, vol. 79, pp. 25 051–25 065, 2020.

[30] L. He, C. Chen, Y. Wang, Q. Fan, C. Chu, J. Xu, and L. Fan, "Network occlusion sensitivity analysis identifies regional contributions to brain age prediction," *bioRxiv*, pp. 2022–10, 2022.

[31] Y. Fu, Y. Huang, Y. Wang, S. Dong, L. Xue, X. Yin, Q. Yang, Y. Shi, and C. Zhuo, "Otfpf: Optimal transport-based feature pyramid fusion network for brain age estimation with 3d overlapped convnext," 2022.

[32] S. Basodi, R. Raja, B. Ray, H. Gazula, A. D. Sarwate, S. Plis, J. Liu, E. Verner, and V. D. Calhoun, "Decentralized brain age estimation using mri data," *Neuroinformatics*, vol. 20, no. 4, pp. 981–990, 2022.

[33] A. Z. Tan, H. Yu, L. zhen Cui, and Q. Yang, "Towards personalized federated learning," *IEEE transactions on neural networks and learning systems*, vol. PP, 2021.

[34] F. Sattler, K.-R. Müller, and W. Samek, "Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, pp. 3710–3722, 2019.

[35] M. Jie, G. Long, T. Zhou, J. Jiang, and C. Zhang, "On the convergence of clustered federated learning," *ArXiv*, vol. abs/2202.06187, 2022.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[37] A. Srinivas, T.-Y. Lin, N. Parmar, J. Shlens, P. Abbeel, and A. Vaswani, "Bottleneck transformers for visual recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 16 519–16 529.

[38] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.