FSLearning: An Efficient Federated Split Learning Framework for Privacy-Preserving Disease Prediction

Bin Li[1], Xiaoqian Jiang[2], Yu-Chun Hsu[2], Arif O. Harmanci[2],
Hongchang Gao[1], and Xinghua Shi[1](✉)

[1] Department of Computer and Information Sciences, College of Science and Technology, Temple University, Philadelphia, PA 19122, USA
mindyshi@temple.edu

[2] D. Bradley McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

Abstract. Federated learning (FL) and split learning (SL) are two prominent distributed learning approaches that enable model training without raw data sharing. SL offers better model privacy than FL by splitting the model architecture between clients and the server, making it preferable for resource-constrained environments. However, SL is slower due to relay-based training across multiple clients. This paper introduces federated split learning, FSLearning, which combines the strengths of both FL and SL, eliminating their inherent drawbacks. FSLearning integrates tensor regression to reduce communication costs and improve training efficiency. Our analysis and empirical results show that FSLearning achieves similar test accuracy and communication efficiency as SL while significantly reducing computation time for multiple clients. Empirical results show that FSLearning reduces transmitted parameters by up to 50% using ResNet3D. By incorporating tensor regression layers (TRLs), FSLearning compresses activations, enabling efficient Homomorphic Encryption (HE) integration. Privacy evaluations confirm that DP achieves the lowest total variation distance (TVD), reducing membership inference risks.

Keywords: Split Learning · Federated Learning · Homomorphic Encryption · Tensor Regression

1 Introduction

Collaborative deep learning methodologies [16], such as federated learning (FL), enable multi-institutional model training without direct data sharing [4,9]. However, traditional deep learning approaches face challenges due to the high dimensionality, scarcity, and privacy-sensitive nature of biomedical datasets. The distributed nature of healthcare further complicates secure data exchange, requiring solutions that balance computational efficiency with rigorous privacy preservation.

R. Bellazzi et al. (Eds.): AIME 2025, LNAI 15734, pp. 222–231, 2025.
https://doi.org/10.1007/978-3-031-95838-0_22

Split learning (SL) [5] has emerged as an alternative to FL, enabling collaborative model training while limiting data exposure to intermediate representations. This is particularly beneficial in privacy-sensitive medical imaging applications, such as MRI-based diagnostics, where leveraging diverse datasets improves model performance while preserving data locality. Unlike FL, which requires frequent exchanges of model weights, SL reduces communication overhead by transmitting only activations and gradients. Whether SL is a sub-class of FL or a distinct paradigm remains debated; following the taxonomy in [7] we treat them as orthogonal, which clarifies how privacy, communication cost, and compute load differ. Prior studies [1,11,12] have demonstrated SL's effectiveness in preserving data privacy using Differential Privacy (DP) across multi-institutional MRI analysis. However, its relay-based training structure often leads to resource underutilization, as only one party interacts with the server at a time. More critically, despite its privacy advantages, SL's communication constraints make it impractical for integrating computationally intensive cryptographic techniques such as Homomorphic Encryption (HE), which provides strong privacy guarantees but introduces significant computational overhead.

Addressing these limitations, this paper introduces FSLearning, a hybrid framework that integrates FL's decentralized coordination with SL's privacy-preserving model partitioning, enabling efficient multi-institutional learning while reducing communication costs. A key innovation of FSLearning is its incorporation of Tensor Regression Layers (TRLs) [8,17], which compress transmitted activations, reducing bandwidth requirements and making it feasible to deploy HE in real-world biomedical applications.

2 Method

The proposed method employs a split learning scheme tailored for collaborative MRI imaging analysis across multiple healthcare institutions, ensuring privacy preservation and data security while enabling the development of robust diagnostic models. The methodology is designed to harness the distributed data without direct sharing of raw MRI images or patient information, thereby addressing the critical concerns of privacy and data security in biomedical research.

2.1 Architecture

Figure 1 illustrates the FSLearning pipeline, which integrates FL and SL to enable collaborative model training while preserving data privacy. Each healthcare institution retains MRI data locally, processing it through the initial layers of a 3D CNN. To reduce communication overhead, a tensor decomposition layer compresses feature representations before transmitting the abstracted activations ("smashed data") to a central server. The server continues the forward pass using deeper CNN layers and computes partial gradients, which are returned to local parties for updating their model segments. This ensures that raw MRI data remains within each institution, while the server lacks access to local model parameters, reinforcing privacy preservation.

Unlike FL, which requires exchanging full model weights, FSLearning minimizes communication by transmitting only activations and gradients. Similar to SL, local parties retain early network layers and, in label-less configurations, maintain final classification layers and labels. The server orchestrates sequential updates across sites, optimizing communication efficiency while maintaining model integrity. This hybrid approach enables scalable and privacy-preserving training across multiple institutions.

For each 3D backbone we tested LeNet3D, AlexNet3D, and ResNet3D, cut the network immediately after the shallow convolutional blocks, push the deeper feature-extractor plus a TRL to the server, and keep the final fully-connected classifier on the local site. This partition leaves the labels and the first 20–40% of parameters inside each hospital while off-loading 60–80% of the FLOPs to the server; the only data crossing the boundary is a single, low-dimensional ($\leq$4096) encrypted activation vector. Each site transmits activations forward and receives gradients back, iterating through mini-batches. The server processes updates sequentially across sites, typically in a randomized order per epoch, optimizing communication efficiency while reducing reliance on full model synchronization.

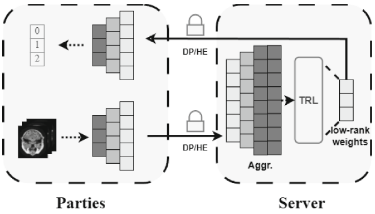

Fig. 1. Overview of the proposed multi-institutional split learning workflow. local parties process raw MRI data through early layers and then send lower-dimensional "smashed data" to the central server. The server refines these representations with additional network layers with a Tensor regression layer, updates its parameters, and returns gradients to each site, preserving data locality and patient privacy.

2.2 Privacy Protection

FSLearning reduces privacy risks by splitting the model so that each healthcare site retains raw patient data and (optionally) the final classification layer, while the central server operates only on intermediate representations. This architectural partition complicates attempts to reconstruct original MRI images or reverse-engineer local models, especially if sufficient dimensionality reduction and non-linear transformations occur early in the network. A label-less configuration further hides class information from the server, ensuring it never has direct access to patient outcomes.

Beyond this structural safeguard, FSLearning can incorporate additional security methods commonly seen in federated learning. For instance, DP can introduce noise to the smashed activations or returned gradients, obscuring individual patient data within the statistics of the model updates. HE libraries (e.g., Zama, SEAL) can also encrypt the intermediate representations while still allowing arithmetic operations on ciphertexts.

A key advantage of FSLearning over traditional FL is its use of a tensor regression layer (TRL) for dimensionality reduction. By compressing the high-volume MRI feature maps into smaller core tensors, the amount of data requiring encryption is substantially reduced, thus mitigating the computational overhead typically associated with HE. Since HE operations grow expensive with data size, lowering the dimensionality of transmissions makes encrypted computations more feasible in large-scale medical applications. Combining TRL-driven compression with label-less split learning can thereby yield a robust, multi-layer privacy defense without excessively sacrificing training efficiency.

3 Experiments

3.1 Datasets

Brain MRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) are used for this study. Specifically, we used 2268 MRI scans from the ADNI3 category to train the LeNet3D, AlexNet3D, and ResNet3D to classify MRI scans for Alzheimer's disease (AD), mild cognitive impairment (MCI), and cognitively normal (CN). Each brain MRI scan is a 3D tensor of intensity values with size 256×256×256. As a result, 1053 scans from CN, 1051 scans from MCI, and 164 scans from AD are selected for this study.

Pre-processing: In our study on fairness modeling using the ADNI III dataset [6], we employed ANTsPy [2,13], the Python interface for Advanced Normalization Tools (ANTs), to standardize and preprocess brain MRI scans. The preprocessing pipeline started with image registration, aligning individual scans to a common anatomical template to ensure consistency across subjects and facilitate comparative analysis. Next, ANTsPy was used for tissue segmentation, delineating distinct brain regions for detailed structural assessment. Cortical thickness was then computed and incorporated as supplementary features to enrich the training data with structural biomarkers relevant to disease progression. To ensure fairness and minimize technical biases, all scans were uniformly processed, preserving data integrity across sites. The dataset was stratified by diagnostic categories (AD, MCI, CN) to ensure equal representation across training and testing sets. To prevent data leakage, we enforced strict subject-level separation, ensuring that no subject's scans appeared in both training and validation folds in cross-validation experiments. This preprocessing framework enhances both the robustness and fairness of our predictive model by mitigating confounding variations and preserving biological relevance.

3.2 Experimental Settings

Experiments are carried out on uniformly distributed and horizontally partitioned image datasets among parties. For quicker experiments and developments, we use the High-Performance Computing (HPC) platform with 4 A5000 GPUs. We run parties and servers on different computing nodes of the cluster provided by HPC.

In our setup, we consider that all participants update the model in each global epoch (i.e., C = 1 during training). We choose ML network architectures and datasets based on their performance and their need to include proportionate participation in our studies. The learning rate is 0.0001 for training models.

3.3 Performance of FSLearning

FSLearning demonstrates notable efficiency improvements over traditional SL and FL by integrating TRL. The inclusion of tensor regression not only compresses feature representations but also enhances computational efficiency, reducing the communication cost per epoch. Additionally, the architecture of FSLearning minimizes the need for full model exchanges, significantly lowering communication overhead compared to FL.

Table 1 provides a side-by-side comparison of FL and FSLearning for LeNet3D, AlexNet3D, and ResNet3D in terms of classification accuracy, total model size (MB), and the number of exchanged parameters per round. For LeNet3D, FL achieves a 0.54 accuracy but requires a large 2680.77MB model and transmits over 435 million parameters each round. In contrast, FSLearning's accuracy is slightly lower at 0.50, yet it drastically reduces both model size (0.11MB) and communication overhead (around 103k parameters). A similar trend holds for AlexNet3D, where FL's higher accuracy of 0.66 comes with 255 million exchanged parameters per round, whereas FSLearning trades a modest drop in accuracy (0.61) for a large reduction in exchanged parameters (46k). Notably, while the overall model size remains the same (609.14MB) for AlexNet3D under both strategies, FSLearning's partial-architecture updates and smashed data transmission still lower the communication volume. Finally, for ResNet3D, FL achieves 0.89 accuracy but transmits more than 32 million parameters per round, whereas FSLearning yields 0.87 accuracy with only 1024 parameters exchanged each round, and both methods use a 1210.84MB model. These results confirm that FSLearning substantially reduces communication costs while retaining competitive accuracy, especially beneficial for large-scale or bandwidth-constrained federated settings.

Effect of Numbers of Users. Figure 2 shows how accuracy and overall training time evolve as the number of parties increases from 1 to 10 in a federated learning setting. We observe that accuracy generally rises with more parties, reflecting the benefit of leveraging a broader data distribution. In particular, the model's accuracy improves from around 0.78 with a single party to approximately

Table 1. Comparison of FL and FSLearning for three 3D CNN architectures, showing classification accuracy (Acc), total model size (in MB), and the number of uploaded parameters (per party per round).

Architecture	FL			FSLearning		
	Acc	Model Size	Uploaded Param.	Acc	Model Size	Uploaded Param.
LeNet3D	0.54	2680.77	435,827,426	0.50	0.11	103,191
AlexNet3D	0.66	609.14	255,812,426	0.61	609.14	46,275
ResNet3D	0.89	1210.84	32,054,402	0.87	1210.84	1,024

0.89 by eight or more parties, indicating a saturation point in performance gains beyond which adding parties offers minimal accuracy improvements.

Conversely, overall training time increases at an accelerating rate due to communication and synchronization overheads that grow with more parties. While smaller increases in the number of parties (e.g., up to three or four) can still benefit from some parallel speedup, subsequent increments see more pronounced jumps in training time as overhead begins to dominate.

These results underscore the trade-off between accuracy gains and training overhead. As the number of parties increases, the computational and communication costs rise faster than the associated improvements in accuracy. Although federated collaboration remains essential for harnessing diverse datasets, we observe that beyond a certain threshold, the overhead begins to eclipse the incremental accuracy benefits. Practitioners should carefully weigh these factors in real-world implementations, ensuring that the added complexity of involving more parties does not undermine the overall efficiency and scalability of the learning process.

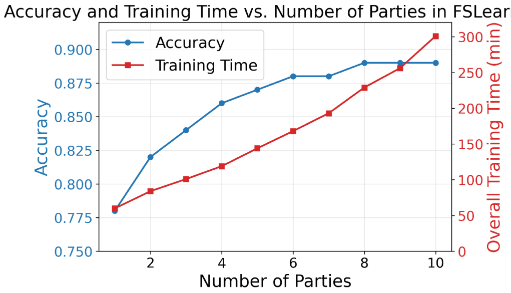

Fig. 2. Accuracy and overall training time as a function of the number of parties in FSLearning. The blue curve represents accuracy and the red curve denotes the overall training time. The results highlight the trade-off between improved model performance and higher training costs in multi-party federated learning settings. (Color figure online)

3.4 Privacy Protection

FSLearning Supports HE and DP. We assessed two privacy mechanisms, HE and DP, inside the FSLearning pipeline, measuring (i) classification accuracy, (ii) resistance to membership-inference attacks (MIA), and (iii) computational/communication overhead. HE keeps every activation encrypted in transit (implemented with the Zama-concrete library [15]); DP injects calibrated Gaussian noise into gradient updates with (ϵ=2.0, δ=10^{-5}).

Table 2 reports ResNet3D results on ADNI III with 95% bootstrap confidence intervals (1000 resamples). DP attains 0.875 accuracy—only 0.6 pp below HE and 0.2 pp below the non-private baseline—while offering the strongest MIA protection, cutting F1/precision/recall scores to 0.42/0.45/0.40. HE preserves the highest accuracy (0.881) but affords markedly weaker MIA defence (F1 = 0.55) and amplifies training time (3.6× vs. baseline) and bandwidth (50 MB vs. 20 MB per round). Because DP achieves near-par accuracy with the lowest privacy-attack scores and halves communication load relative to HE, we regard DP-enhanced FSLearning as the more practical choice for real-world, multi-site biomedical deployments, whereas HE remains advantageous in settings that mandate end-to-end encrypted transport—a practicality enabled by the tensor-regression layer's substantial parameter compression. Furthermore, Table 3 shows that FSLearning with DP excels on the majority MCI class (recall 0.86, precision 0.84) and maintains high performance on CN (0.92, 0.93), while still achieving a respectable 0.68 recall for the minority AD class. This result indicates that the framework boosts early-stage (MCI) detection while maintaining acceptable accuracy for the much rarer AD class, underscoring its utility for clinical screening under class-imbalanced conditions.

Table 2. ResNet3D on ADNI: privacy mechanisms versus the unprotected baseline. Values in brackets are 95 % bootstrap CIs (N=1000).

Metric	HE	DP	Baseline
Accuracy ↑	0.881 [0.870, 0.892]	0.875 [0.864, 0.887]	0.877 [0.866, 0.889]
F1 (MIA) ↓	0.55 [0.51, 0.59]	**0.42** [0.38, 0.46]	0.68 [0.63, 0.72]
Precision (MIA) ↓	0.58 [0.54, 0.62]	**0.45** [0.40, 0.49]	0.75 [0.70, 0.78]
Recall (MIA)↓	0.53 [0.49, 0.58]	**0.40** [0.36, 0.44]	0.62 [0.57, 0.66]
Training time (s/epoch) ↓	320	130	**90**
Communication (MB/round) ↓	50	25	**20**

FSLearning Defends Against MIA. MIA attempt to decide whether a particular sample participated in training, posing serious risks when re-identification would violate medical confidentiality. FSLearning lowers that risk by keeping raw data on-site and exchanging only compressed or encrypted activations. When layered with DP or HE this partitioning further obscures clues an adversary

Table 3. Per-class recall and F1 of FSLearning with DP on ADNI (95 % bootstrap CIs, N = 1000).

Metric	AD	MCI	CN
Recall	0.68 [0.63, 0.73]	0.86 [0.83, 0.89]	0.92 [0.89, 0.95]
Precision	0.66 [0.61, 0.71]	0.84 [0.80, 0.88]	0.93 [0.90, 0.95]
F1	0.67 [0.62, 0.72]	0.85 [0.82, 0.88]	0.93 [0.91, 0.95]
AUC	0.90 [0.88, 0.92]	0.95 [0.93, 0.97]	0.97 [0.95, 0.98]

could exploit. Figure 3 illustrates train (member) versus holdout (non-member) accuracy distributions under three privacy settings (Baseline, DP, and HE) for different numbers of parties ($P = 5, 7, 9$). Each subplot shows histograms of the accuracy achieved on train and holdout samples, along with the corresponding total variation distance (TVD) [3, 14].

In the baseline row, train accuracies are clearly shifted higher than holdout, yielding large total-variation distances (TVD: 106.12 at P=5; 132.36 at P=9), evidence that an attacker could reliably infer membership. With DP (middle row) calibrated noise blurs individual contributions, shrinking the gap between

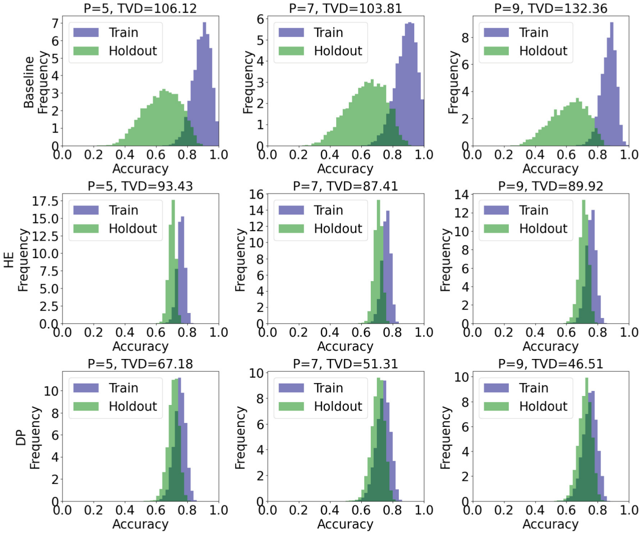

Fig. 3. Membership inference attack (MIA) results on a multi-institutional MRI dataset under three privacy settings: Baseline, DP, and HE. Each subplot compares the train (member) and holdout (non-member) accuracy distributions for $P = 5$, $P = 7$, and $P = 9$ parties, where P denotes the number of participating institutions. The total variation distance (TVD) reported in each subplot quantifies how separable the two distributions are, with higher TVD indicating greater vulnerability to membership inference.

train and holdout. TVD drops to 67.18 at P=5 and 46.51 at P=9, the lowest values across all settings. The broader train histogram reflects the randomness injected into gradients, and accuracy declines only modestly (see Table 2). In the HE setting (bottom row), raw activations remain encrypted in transit, so an attacker cannot observe them directly; however, once decrypted on the server they still contain detailed signal, giving a TVD (e.g., 93.43 at P=5, 89.92 at P=9) that lies between DP and the baseline. Thus HE improves privacy relative to no protection but is less effective than DP for mitigating MIA.

Finally, increasing the number of parties P without privacy (baseline) tends to amplify overfitting and raise TVD, while both DP and HE benefit from additional parties, further narrowing the train-holdout gap. Overall, these results confirm that privacy-enhancing techniques, especially differential privacy, meaningfully reduce membership-inference risk, and that wider collaboration can enhance generalization under secure training protocols.

4 Conclusion

This paper introduced FSLearning, a Federated Split Learning framework that integrates tensor regression to improve communication efficiency while ensuring privacy preservation. By combining FL and SL, FSLearning mitigates FL's high communication costs and SL's computational inefficiencies, offering a scalable and efficient solution for multi-client settings. Experimental results demonstrated that FSLearning achieves accuracy comparable to FL while significantly reducing communication overhead by up to 50%. Furthermore, the TRL-induced 90% reduction in transmitted activation size makes HE feasible on commodity GPUs.

However, challenges remain, such as the computational overhead introduced by HE, which may limit its real-time applicability, and potential trade-offs in model complexity due to tensor regression compression. Addressing these challenges is key to optimizing the framework for broader real-world deployment.

Future work will explore adaptive tensor compression [10] to further enhance efficiency and hybrid privacy mechanisms that dynamically switch between HE and DP based on computational constraints. Additionally, expanding FSLearning to multi-modal biomedical datasets and integrating it into real-world clinical applications will be critical to validating its effectiveness in broader healthcare domains.

Acknowledgements. XJ is CPRIT Scholar in Cancer Research (RR180012), and he was supported in part by the Christopher Sarofim Family Professorship, UT Stars award, UTHealth startup, the National Institute of Health (NIH) under award number R01AG066749, R01AG066749-03S1, R01LM013712, U24LM013755 and U54HG012510, and the National Science Foundation (NSF) #2124789. This manuscript was also partially supported by an NIH grant to XS (1R01NS140142-01).

Data Availability Statement. The datasets analysed during the current study are available in the Alzheimer's Disease Neuroimaging Initiative (ADNI) repository, https://adni.loni.usc.edu/.

References

1. Abedi, A., Khan, S.S.: FEDSL: federated split learning on distributed sequential data in recurrent neural networks. Multimedia Tools Appl. **83**(10), 28891–28911 (2024)
2. Avants, B.B., Tustison, N., Song, G., et al.: Advanced normalization tools (ANTS). Insight J. **2**(365), 1–35 (2009)
3. Chen, J., et al.: Discriminative forests improve generative diversity for generative adversarial networks. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 11338–11345 (2024)
4. Fang, H., Qian, Q.: Privacy preserving machine learning with homomorphic encryption and federated learning. Future Internet **13**(4), 94 (2021)
5. Hu, Z., Zhou, T., Wu, B., Chen, C., Wang, Y.: A review and experimental evaluation on split learning. Future Internet **17**(2), 87 (2025)
6. Jack Jr, C.R., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magnetic Resonance Imag. Official J. Int. Soc. Magnetic Resonance Med. **27**(4), 685–691 (2008)
7. Kairouz, P., et al.: Advances and open problems in federated learning. Found. Trends® Mach. Learn. **14**(1–2), 1–210 (2021)
8. Kossaifi, J., Panagakis, Y., Anandkumar, A., Pantic, M.: Tensorly: tensor learning in python. J. Mach. Learn. Res. **20**(26), 1–6 (2019). http://jmlr.org/papers/v20/18-277.html
9. Li, B., Gao, H., Shi, X.: FedDP: secure federated learning with differential privacy for disease prediction. In: International Conference on Computational Advances in Bio and Medical Sciences, pp. 119–131. Springer (2023). https://doi.org/10.1007/978-3-031-82768-6_11
10. Nie, C., Wang, H., Tian, L.: Adaptive tensor networks decomposition. In: BMVC, p. 148 (2021)
11. Poirot, M.G.: Split learning in health care: multi-center deep learning without sharing patient data. Master's thesis, University of Twente (2020)
12. Thapa, C., Arachchige, P.C.M., Camtepe, S., Sun, L.: SPLITFED: when federated learning meets split learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 8485–8493 (2022)
13. Tustison, N.J., et al.: Longitudinal mapping of cortical thickness measurements: an Alzheimer's disease neuroimaging initiative-based evaluation study. J. Alzheimers Dis. **71**(1), 165–183 (2019)
14. Verdú, S.: Total variation distance and the distribution of relative information. In: 2014 Information Theory and Applications Workshop (ITA), pp. 1–3. IEEE (2014)
15. Zama: Concrete: TFHE Compiler that converts Python programs into FHE equivalent (2022). https://github.com/zama-ai/concrete
16. Zhang, D., Chen, X., Wang, D., Shi, J.: A survey on collaborative deep learning and privacy-preserving. In: 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), pp. 652–658. IEEE (2018)
17. Zhou, H., Li, L., Zhu, H.: Tensor regression with applications in neuroimaging data analysis. J. Am. Stat. Assoc. **108**(502), 540–552 (2013)