Privacy-Preserving Dysphonia Detection based on Distributed Deep Learning

Jiarui Xu[1], Zhuohang Li[2], Yike Li, MD, PhD[3], Zhiyu Wan, PhD[1],

[1]School of Biomedical Engineering, ShanghaiTech University, Shanghai, China; [2]Vanderbilt University, Nashville, TN, USA; [3]Vanderbilt University Medical Center, Nashville, TN, USA

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY.

Introduction

Auditory-perceptual assessment remains to be the gold standard for evaluating instances of voice diseases, yet its value is bounded by the subjectivity of raters and the limited accessibility of consultable materials within reach.

Value of Deep Learning: Advanced techniques such as deep learning hold profound value for improving clinical decision-support and therapeutic efficiency.

Privacy and Data Silos: Due to concerns regarding patients' privacy as well as the inherited nature of data silos between healthcare entities, data sharing is significantly restricted in practice.

Distributed Learning as solution: distributed learning enables collaborative model training without sharing raw data, among which federated learning (FL) and split learning (SL) are two widely adopted approaches.

Method

Dataset: Source & Cohort: Retrospectively derived from routine clinical voice recordings at Eye, Ear, Nose and Throat Hospital, Fudan University.

- Included 605 adult native Chinese (Mandarin) speakers (235 healthy, 370 patients with diagnoses like polyps, glottal incompetence, etc.).
- Sustained vowel recordings are of /a/ and /i/ (44.1kHz sampling rate, 16-bit resolution).

Preprocessing: Silent intervals removed; segments trimmed to avoid fade effects; remaining voice data uniformly split into 1.5s non-overlapping audio clips. Mel-spectrogram (128 filter banks) plus its first and second derivatives (constituting a complete three-channel time-frequency feature vector).

Distributed Learning Framework: We adopted SplitFedv1 (SFLV1), a horizontal Split Learning variant, to achieve privacy-preserving dysphonia detection. This framework integrates parallel client computation and dual-end federated averaging.

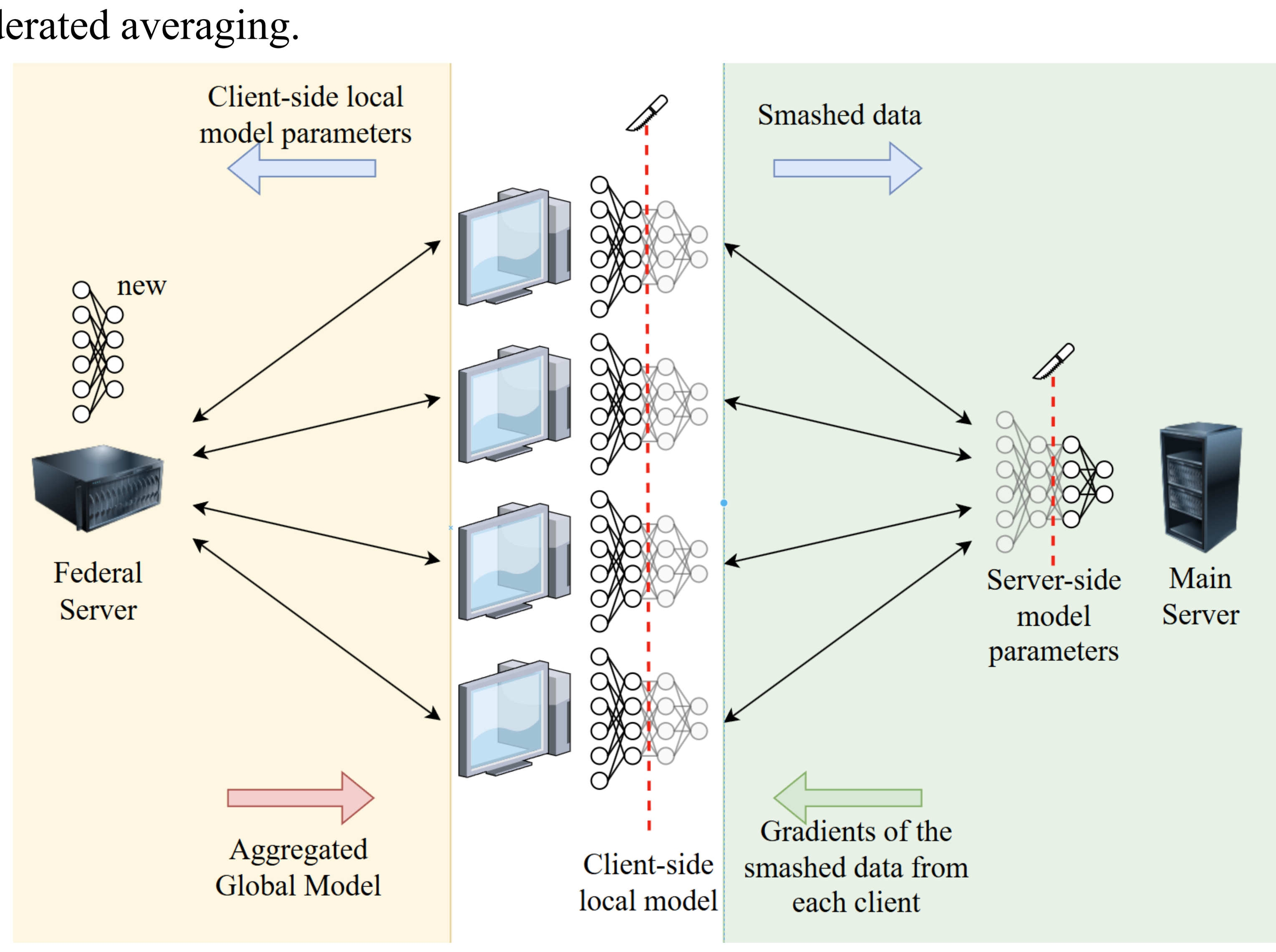

Figure 1. Schematic of the SplitFedv1 (SFLV1) Architecture.

Conclusion

- Proposed a privacy-preserving dysphonia detection system using **distributed deep learning (FL and SL)**.
- Adapted the **Audio Spectrogram Transformer** to the SL framework, validating its effectiveness and structural extensibility across various neural network architectures.
- Future research will focus on optimizing split strategies for advanced architectures and exploring the **privacy-utility trade-offs**.

Reference

1. Thapa C, et al. Splitfed: When federated learning meets split learning. *AAAI* 2022 (Vol. 36, No. 8, pp. 8485-8493).
2. Li Z, Yan C, Zhang X, Gharibi G, Yin Z, Jiang X, Malin BA. Split learning for distributed collaborative training of deep learning models in health informatics. InAMIA Annual Symposium Proceedings 2024 Jan 11 (Vol. 2023, p. 1047).
3. Chen Z, Zhu P, Qiu W, Guo J, Li Y. Deep learning in automatic detection of dysphonia: Comparing acoustic features and developing a generalizable framework. *Int. J. Lang. Commun. Disord.* 2023 Mar;58(2):279-94.

Result

Key Performance Findings:

- A basic convolutional neural network (CNN) was used to evaluate diagnostic performance of distinct distributed frameworks.
- SFLV1 demonstrates faster convergence and higher final test accuracy compared to federated learning.
- We got comparative average test accuracies of around **91% (CL), 82% (FL), 84% (SFLV1 with CNN)** at 200 global epochs.
- Auxiliary metrics comprising AUROC, AUPRC, F1 score, and Kappa coefficient were employed, convergence characteristics were highly consistent with that of test accuracy.

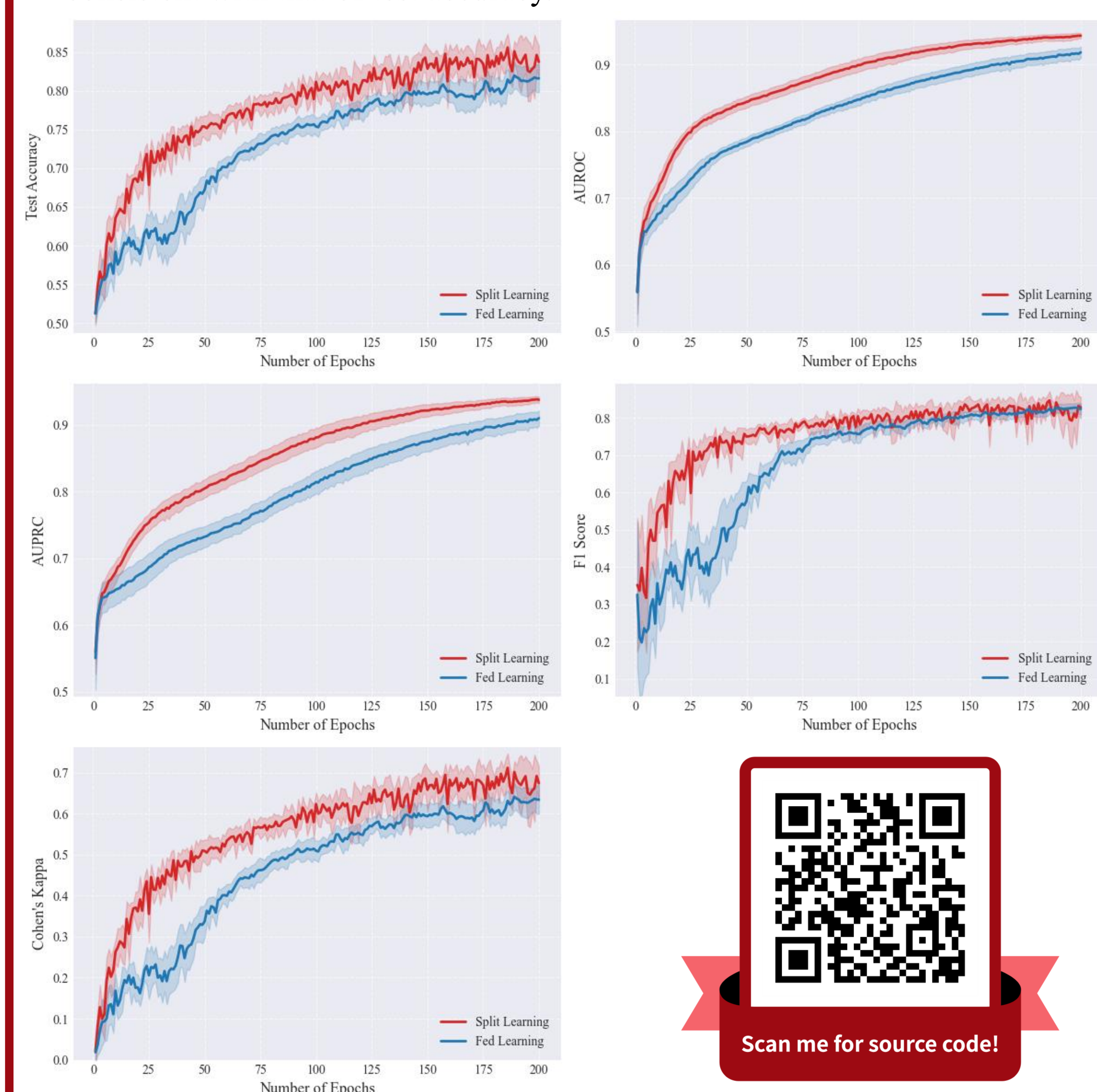

Figure 2. Test accuracy, AUROC, AUPRC, F1 score, and Kappa coefficient of federated learning and SFLV1.

Advanced Model and Extensibility

- The centralized learning (CL) model was incorporated as a key control group and overwhelmingly outperformed both distributed learning frameworks.
- We modified the **Audio Spectrogram Transformer (AST)** to be compatible with two distributed frameworks and the SFLV1 framework reached **91% test accuracy**.

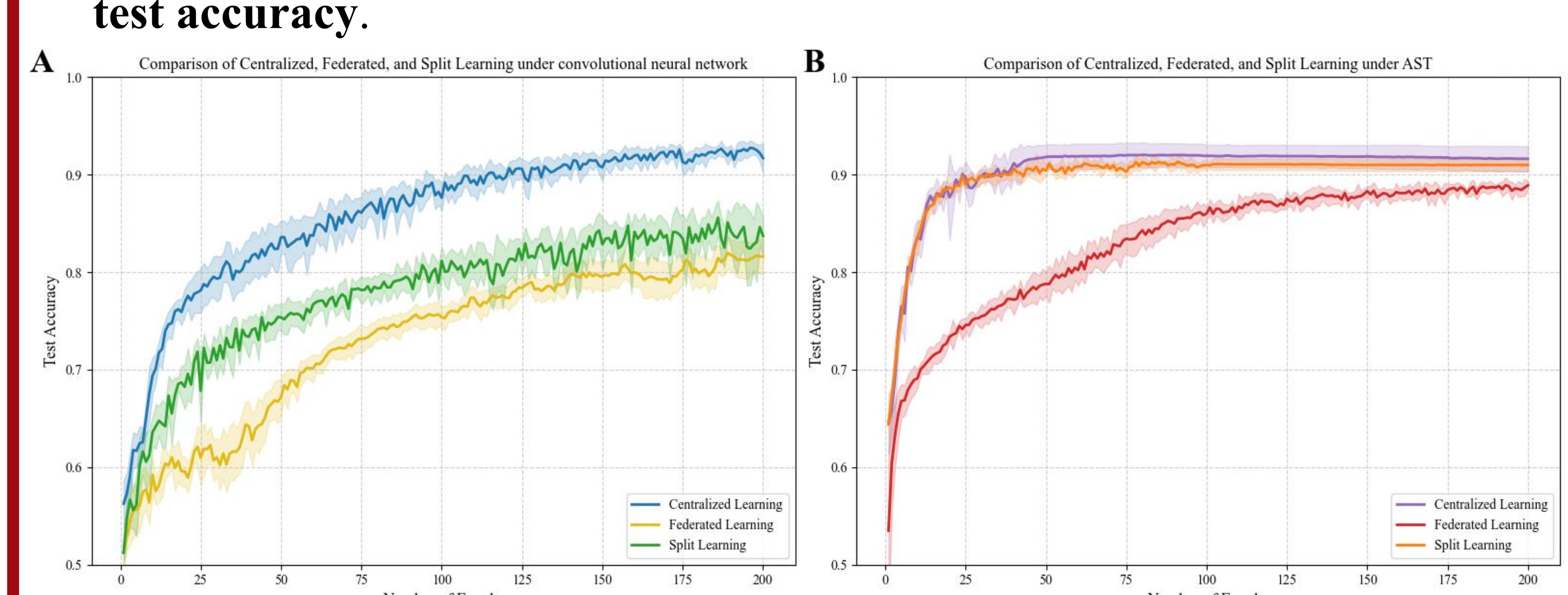

Figure 3. Test accuracy of different learning frameworks. AST, Audio Spectrogram Transformer.

Privacy Enhancement

Beyond the inherent structural isolation advantage by SL, we further applied Differential Privacy (DP)-based gradient perturbation on the client-side model in an attempt to enhance overall defense capability.

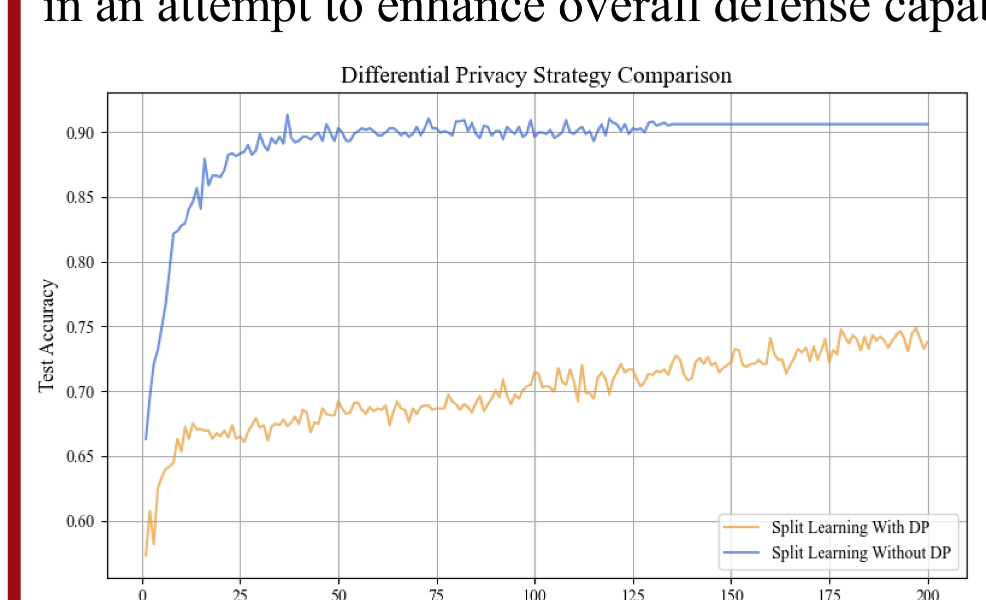

Figure 4. The impact of client-side DP mechanism on the convergence of split learning.