Multi-center Weakly Supervised Computational Pathology on Whole Slide Images using Federated Learning

HARVARD MEDICAL SCHOOL

BRIGHAM AND WOMEN'S HOSPITAL

Dana-Farber Cancer Institute

Mahmood Lab AI for Pathology

Ming Y. Lu, Dehan Kong, Jana Lipkova, Richard J. Chen, Rajendra Singh, Drew F. K. Williamson, Tiffany Y. Chen, Faisal Mahmood

mlu16@bwh.harvard.edu - faisalmahmood@bwh.harvard.edu

www.mahmoodlab.org

Abstract

We present a federated learning framework for both weakly-supervised classification and survival prediction on whole slide images and validate our approach on three different computational pathology tasks, using multi-institutional data from the TCGA and Brigham and Women's Hospital. Our approach is privacy preserving, does not require manual annotations during training, is interpretable and can be applied at scale.

Motivation

- Accurate, robust AI predictive models often require large amounts of diverse, high-quality labeled data, which can be difficult or impossible to curate at a single institution.
- Federated learning allows machine learning models to be developed on multi-institutional data without requiring direct data sharing, alleviating the burden of navigating around many challenges of sharing sensitive patient data and large cohorts of histopathology images at the tera- or peta-byte scale.
- There is a need for the development and validation of federated learning algorithms for a wide range of computational pathology tasks.

Weakly-supervised learning on WSIs

First, the tissue regions of each WSI are segmented and divided into patches and encoded by a pretrained CNN into low-dimensional embeddings $\{\boldsymbol{z}_m\}$, $\boldsymbol{z}_m \in \mathbb{R}^{1024}$. This results in each WSI being described by a matrix of patch embeddings: $\boldsymbol{H}_j \in \mathbb{R}^{M_j \times 1024}$, where M_j is the number of patches in the j^{th} WSI. After projecting $\{\boldsymbol{z}_{j,m}\}$ into $\{\boldsymbol{h}_{j,m}\}$, $\boldsymbol{h}_{j,m} \in \mathbb{R}^{512}$, the contribution of each patch $a_{j,m}$ is given by:

$$a_{j,m} = \frac{\exp\left\{\mathbf{W}_a\left(\tanh\left(\mathbf{V}_a\mathbf{h}_{j,m}^\top\right) \odot \operatorname{sigm}\left(\mathbf{U}_a\mathbf{h}_{j,m}^\top\right)\right)\right\}}{\sum_{m=1}^{M_j}\exp\left\{\mathbf{W}_a\left(\tanh\left(\mathbf{V}_a\mathbf{h}_{j,m}^\top\right) \odot \operatorname{sigm}\left(\mathbf{U}_a\mathbf{h}_{j,m}^\top\right)\right)\right\}}$$

where $\boldsymbol{W}_a$, $\boldsymbol{U}_a$ and $\boldsymbol{V}_a$ are learnable network parameters. Using weighted average pooling, the WSI is in turn described by $\boldsymbol{h}_{bag_j} \in \mathbb{R}^{512}$:

$$\mathbf{h}_{bag_j} = \textbf{Attn-pool}(\mathbf{A}_j, \mathbf{H}_j) = \sum_{m=1}^{M_j} a_{j,m}\mathbf{h}_{j,m}$$

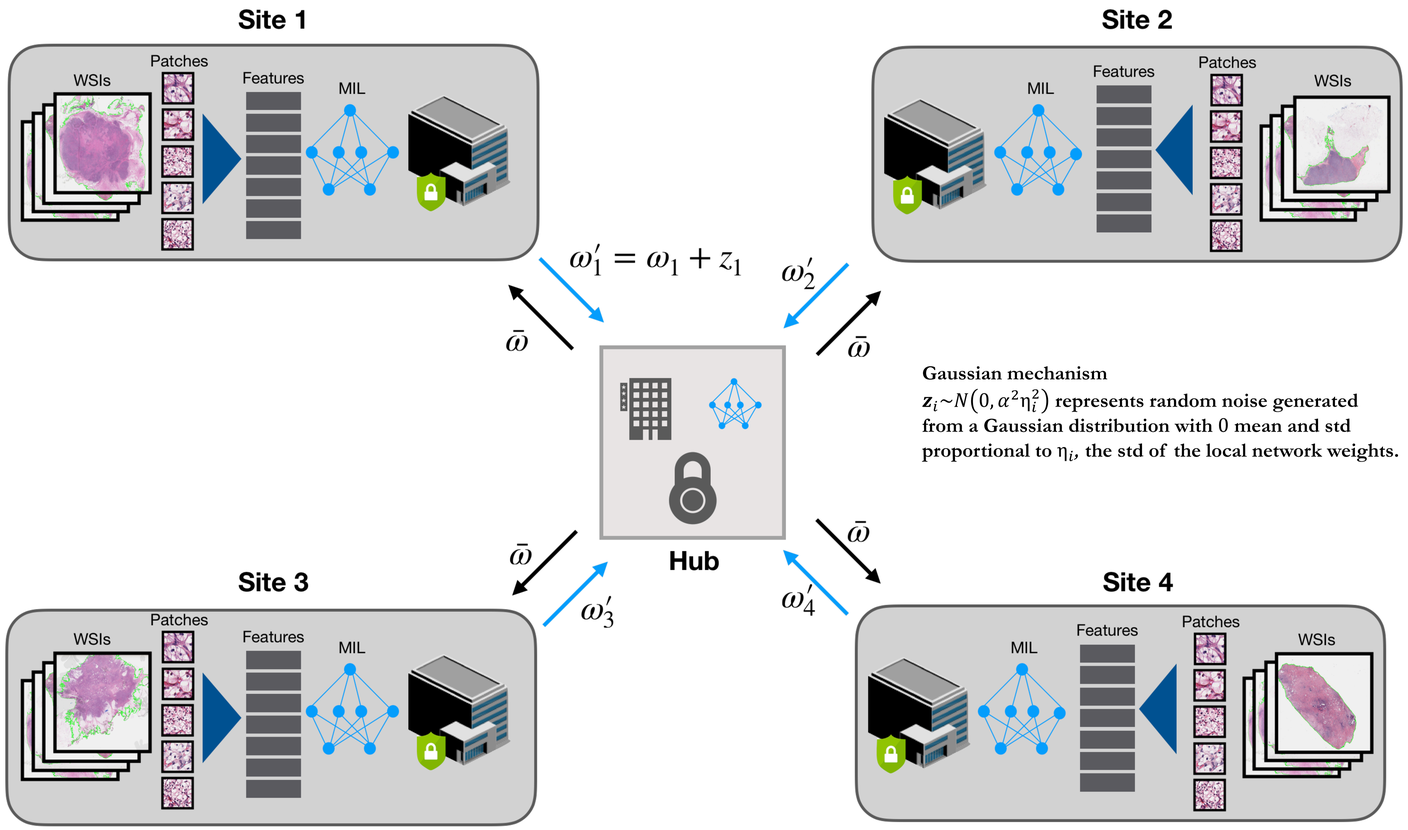
Depending on the task, the last layer of the network either predicts class-probability scores (classification) or models the hazard function (survival prediction):

$$f_{hazard}(r \mid \mathbf{h}_{bag_j}) = P(T_j = r \mid T_j \geq r, \mathbf{h}_{bag_j})$$

For classification, the network is supervised via the cross-entropy loss, and for survival, we used the nll loss given below, where f_{surv} is the survival function, c_j the censorship status and Y_j the ground truth discrete follow-up time:

$$\begin{aligned} L = -l = &-c_j \cdot \log\left(f_{surv}\left(Y_j \mid \mathbf{h}_{bag_j}\right)\right) \\ &- (1-c_j) \cdot \log\left(f_{surv}\left(Y_j - 1 \mid \mathbf{h}_{bag_j}\right)\right) \\ &- (1-c_j) \cdot \log\left(f_{hazard}\left(Y_j \mid \mathbf{h}_{bag_j}\right)\right) \end{aligned}$$

Federated Learning for Weakly-supervised Classification and Survival Prediction on Whole Slide Images



Gaussian mechanism

$z_i \sim N(0, \alpha^2\eta_i^2)$ represents random noise generated from a Gaussian distribution with 0 mean and std proportional to η_i, the std of the local network weights.

Experimental Setup

- Classification tasks:
  - Renal cell carcinoma (RCC) histologic subtyping into clear cell (CCRCC), papillary (PRCC) and chromophobe (CHRCC)
  - Breast invasive carcinoma (BRCA) histologic subtyping into ductal (IDC) and lobular (ILC)
- Survival prediction:
  - CCRCC survival prediction and risk-based patient stratification
- Datasets:
  - TCGA-KIRC (519 WSIs), TCGA-KIRP (297 WSIs), TCGA-KICH (121 WSIs), TCGA-BRCA (1056 WSIs), BWH In-house (247 RCC WSIs and 1070 BRCA WSIs)
  - TCGA datasets are separated into multiple groups of institutions based on tissue source sites
- Learning algorithms:
  - Privacy-preserving federated learning using federated averaging + Gaussian mechanism with adjustable hyperparameter α
  - Attention multiple instance learning based classification (cross-entropy loss) and survival prediction (negative log-likelihood loss for discrete survival model).
  - Adam optimizer: 2e-4 lr, 1e-5 weight-decay
- Evaluation:
  - 70/15/15 train, val, test partitions
  - AUC, Error, bACC (balanced accuracy), F1, mAP, Cohen's κ
  - C-index, log-rank test

Results

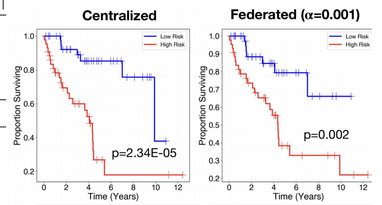
RCC Subtyping

	AUC (95% CI) ↑	Error ↓	bAcc ↑	F1 ↑	mAP ↑	Cohen's κ ↑
Site 1 only	0.975 (0.964 - 0.986)	0.112	0.850	0.888	0.955	0.787
Site 2 only	0.964 (0.948 - 0.980)	0.165	0.821	0.835	0.941	0.699
Site 3 only	0.951 (0.933 - 0.969)	0.176	0.811	0.824	0.917	0.683
Site 4 only	0.957 (0.941 - 0.974)	0.153	0.828	0.847	0.930	0.715
Centralized	0.988 (0.981 - 0.995)	0.071	0.950	0.929	0.980	0.871
Federated, $\alpha = 0.001$	0.988 (0.981 - 0.995)	0.071	0.931	0.929	0.978	0.870
Federated, $\alpha = 0.01$	0.986 (0.978 - 0.994)	0.088	0.922	0.912	0.975	0.839
Federated, $\alpha = 0.1$	0.986 (0.979 - 0.994)	0.071	0.917	0.929	0.976	0.869
Federated, $\alpha = 1$	0.685 (0.638 - 0.732)	0.576	0.416	0.424	0.539	0.129

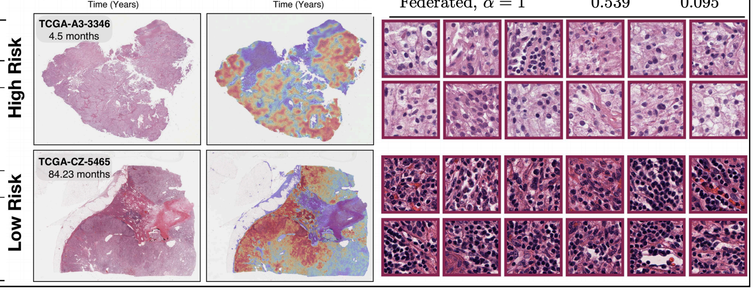
BRCA Subtyping

	AUC (95% CI) ↑	Error ↓	bAcc ↑	F1 ↑	mAP ↑	Cohen's κ ↑
Site 1 only	0.866 (0.808 - 0.923)	0.143	0.720	0.558	0.667	0.473
Site 2 only	0.853 (0.795 - 0.912)	0.121	0.782	0.649	0.618	0.575
Site 3 only	0.800 (0.719 - 0.881)	0.128	0.668	0.494	0.605	0.434
Site 4 only	0.874 (0.821 - 0.928)	0.106	0.750	0.638	0.750	0.580
Centralized	0.946 (0.910 - 0.981)	0.078	0.815	0.747	0.874	0.703
Federated, $\alpha = 0.001$	0.928 (0.886 - 0.970)	0.100	0.836	0.724	0.837	0.663
Federated, $\alpha = 0.01$	0.927 (0.885 - 0.969)	0.100	0.836	0.724	0.834	0.663
Federated, $\alpha = 0.1$	0.929 (0.886 - 0.972)	0.097	0.838	0.730	0.840	0.672
Federated, $\alpha = 1$	0.650 (0.569 - 0.732)	0.174	0.564	0.243	0.341	0.176

Survival Prediction



	c-index ↑	p-value
Site 1 only	0.532	0.115
Site 2 only	0.536	0.141
Site 3 only	0.713	1.937e-04
Centralized	0.715	2.336e-05
Federated, $\alpha = 0.001$	0.713	1.847e-03
Federated, $\alpha = 0.01$	0.690	8.121e-03
Federated, $\alpha = 0.1$	0.700	0.012
Federated, $\alpha = 1$	0.539	0.095



Discussion

- We generalize attention-based multiple instance learning to weakly-supervised federated learning for multi-institutional computational pathology on WSIs, requiring only slide-level/patient-level labels
- Our approach is privacy-preserving, interpretable, generally applicable to both classification and survival prediction and can easily be used on large WSI datasets

Poster

Lab Website