

Attribute-Aware Federated Learning with Pseudo-Bag Augmentation for Multi-Center WSI Classification

Yingjiao Deng¹, Qing Zhang¹, Hui Tang³, Chunhua Zhou², Jiansheng Wang¹, Li Sun¹, Yan Wang¹, Duowu Zou², Hongkai Xiong¹, and Qingli Li^{1*}

¹ Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai, 200241, China

² Department of Gastroenterology, Rujin Hospital, School of Medicine, Shanghai JiaoTong University, Shanghai, China

³ The First People Hospital of Yunnan Province, Yunnan, China

Abstract. Federated learning (FL) on whole slide images (WSIs) remains challenged by severe cross-center heterogeneity arising from inconsistent staining protocols, scanning optics, and slide digitization settings. Existing FL approaches typically rely on coarse feature statistics or require iterative transmission of synthetic images, making it difficult to capture domain-specific pathology priors while preserving privacy. To solve these problems, we propose FedGMM-Bag, a generative federated learning framework that performs cross-center distribution alignment through attribute-aware pseudo-bag generation without sharing raw WSIs. Each center constructs a compact multi-attribute Gaussian Mixture Model (GMM) group of its local data, which is communicated once and used to synthesize cross-domain pseudo-bags for augmenting local MIL training. FedGMM-Bag introduces two complementary mechanisms: distribution replay to introduce external morphological diversity and covariance-aligned feature mapping to achieve fine-grained cross-domain alignment. Comprehensive experiments on three heterogeneous multi-center datasets demonstrate that FedGMM-Bag consistently outperforms existing federated and pseudo-bag methods on both seen and unseen centers. These results demonstrate its strong cross-domain generalization and its suitability for privacy-preserving federated pathology under real-world heterogeneity.

Keywords: Federated Learning · Gaussian Mixture Model · Multi-Instance Learning · Pathology · Whole Slide Image

1 Introduction

Whole-slide images (WSIs) offer detailed visualizations of tissue morphology and are widely used in computer-aided diagnostic systems for cancer diagnosis, prognostic assessment, and treatment planning [8, 14]. Owing to the absence

* Corresponding author. Email: qlli@cs.ecnu.edu.cn

of patch-level supervision, slide-level analysis is predominantly modeled using multiple instance learning (MIL) [13, 4, 10]. WSIs collected in clinical practice inherently exhibit substantial variability due to differences in staining protocols, scanner hardware, tissue preparation, and other site-specific factors. Such heterogeneity arises even within a single institution and is further amplified across centers. To build clinically robust models that generalize to this diversity, multi-center learning is essential, but strict privacy requirements prohibit raw WSI exchange among institutions, making centralized learning infeasible. Federated learning (FL) [15, 16, 22, 3] offers a principled paradigm for collaborative training without sharing raw data. Existing FL methods mainly focus on parameter averaging [15, 16], proximal regularization [22], feature alignment [9], and domain generalization [20], and have shown promise in privacy-preserving medical image analysis [5, 24]. However, directly applying these techniques to WSIs is non-trivial, as they overlook the instance- and bag-level feature structures of WSIs and thus struggle with the pronounced cross-center heterogeneity.

Recent federated WSI studies explore three main directions: MIL-based collaborative training [12], stain-style harmonization [19], and synthetic-slide distillation [7]. While effective to some extent, these strategies either do not correct feature-space domain shifts, rely on low-dimensional color statistics, or require computationally expensive image generation without modeling fine-grained intra-bag variability. This highlights the need for a compact, distribution-aware representation that captures center-specific WSI characteristics. In centralized WSI classification, pseudo-bag generation and instance mining have proven effective for expanding training diversity and improving MIL robustness [11, 23]. Gaussian mixture models (GMMs) [21, 17] offer a principled way to capture feature densities and provide lightweight, compact statistical descriptors suitable for federated communication. These observations motivate a generative, GMM-guided pseudo-bag strategy for addressing cross-center heterogeneity in federated WSI analysis.

In this paper, we propose **FedGMM-Bag**, an attribute-aware generative federated framework for WSI classification. Instead of relying on raw image sharing, low-order stain statistics, or computationally expensive synthetic slides, each participating center communicates a compact multi-attribute Gaussian mixture representation that summarizes the essential characteristics of its WSIs while preserving privacy. This representation enables cross-domain pseudo-bag synthesis that enriches local MIL training with external domain cues without exposing raw patches or high-dimensional features. By combining distribution replay and covariance-aligned feature mapping, FedGMM-Bag reduces non-i.i.d. variation across centers and substantially strengthens cross-center generalization with very limited communication overhead. Experiments on our in-house PANC dataset and two public datasets demonstrate the robustness of FedGMM-Bag.

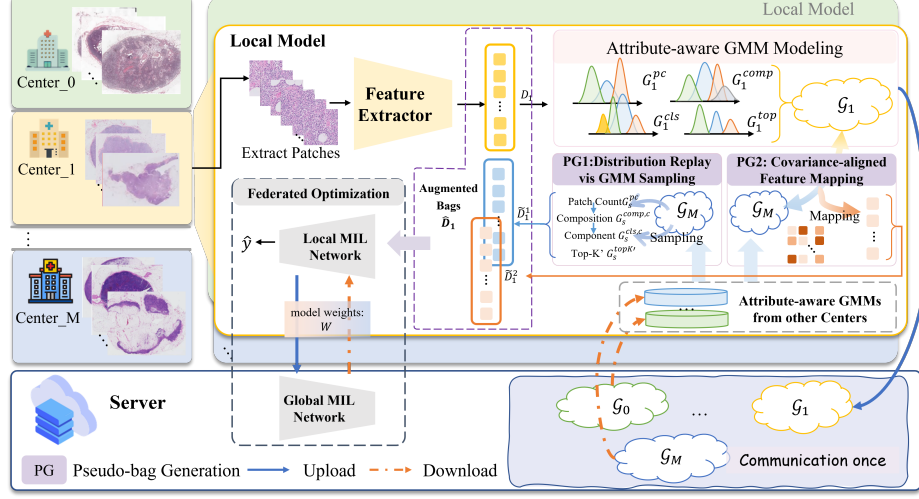


Fig. 1. Overview of the proposed FedGMM-Bag framework, which models local WSI statistics using attribute-aware GMM groups and synthesizes pseudo-bags through distribution replay and covariance-aligned mapping for federated optimization.

2 Method

2.1 Federated MIL Preliminary

Given M participating centers in a federated MIL setting, the goal is to collaboratively train a global WSI classifier that performs robustly under heterogeneous, non-i.i.d. data distributions while preserving patient privacy. Center $m \in \{1, \dots, M\}$ holds a local dataset $D_m = \{(X_i^m, y_i^m) \mid i = 1, \dots, N_m\}$, where X_i^m denotes a WSI containing n_i^m patches and $y_i^m \in \{1, \dots, C\}$ is the slide-level label. Each WSI X_i^m is tiled into patches $\{x_{i,j}^m\}_{j=1}^{n_i^m}$, and a pretrained feature extractor Φ maps each patch to a d -dimensional patch feature vector $z_{i,j}^m = \Phi(x_{i,j}^m) \in \mathbb{R}^d$, yielding a bag-level representation $Z_i^m = \{z_{i,j}^m\}_{j=1}^{n_i^m}$. In federated learning, each center maintains a local MIL model $f_m(\cdot; \theta_m)$ that aggregates instance features within Z_i^m into a bag-level prediction $\hat{y}_i^m = f_m(Z_i^m; \theta_m)$. The server periodically aggregates local model $\{\theta_m\}_{m=1}^M$ into a global model $f(\cdot; \theta)$.

To address cross-center domain heterogeneity in federated WSI classification, we propose FedGMM-Bag, a generative federated learning framework for robust multi-center MIL. The overall framework is illustrated in Fig. 1. Each center is summarized by an attribute-aware GMM group, and these compact generative descriptors are used to synthesize cross-domain pseudo-bags that augment local MIL optimization without exposing raw WSIs or high-dimensional features. Using both real and synthesized bags, each center performs federated MIL training to update its local model and contribute to the global aggregation.

2.2 Attribute-Aware Gaussian Mixture Models

To capture heterogeneous statistical attributes of WSIs at center m , an attribute-aware GMM group of center m is constructed:

$$\mathcal{G}_m = \{\mathcal{G}_m^{pc}, \mathcal{G}_m^{cls}, \mathcal{G}_m^{comp}, \mathcal{G}_m^{top}\}, \quad (1)$$

where $\mathcal{G}_m^{pc}$ characterizes patch-count variability across slides, $\mathcal{G}_m^{cls}$ models class-wise feature densities, $\mathcal{G}_m^{comp}$ represents intra-bag composition patterns, and $\mathcal{G}_m^{top}$ summarizes the distributions of top- K' discriminative features. Together, $\mathcal{G}_m$ forms a compact, privacy-preserving summary of the local WSI domain.

Patch-Count Attribute Modeling. The number of patches per slide varies substantially across WSIs and centers. Let n denote the patch count of a WSI. A log-transform $\log n$ is used when estimating the patch-count GMM for numerical stability. Center m models the log-count distribution by a 1-D GMM:

$$p(\log n) = \sum_{r=1}^{K_{pc}} \alpha_r \mathcal{N}(\log n \mid \mu_r^n, \sigma_r^2), \quad (2)$$

where K_{pc} is the number of components and $\{\alpha_r, \mu_r^{(n)}, \sigma_r^{(n)}\}$ are mixture weights, means and standard deviations, respectively.

Class-Specific Attribute Modeling. For each class c , patch features belonging to slides of class c at center m are modeled by a class-conditional GMM:

$$p(z \mid y = c) = \sum_{r=1}^{K_{cls}} \gamma_{c,r} \mathcal{N}(z \mid \mu_{c,r}, \Sigma_{c,r}), \quad (3)$$

where K_{cls} denotes the number of mixture components per class, and $\{\gamma_{c,r}, \mu_{c,r}, \Sigma_{c,r}\}$ are the mixing coefficients, means and diagonal covariance matrices. These components capture multi-modal morphological patterns within each category.

Bag-Composition Attribute Modeling. Within a WSI of class c , each patch feature is assigned to a component of G_m^{cls} based on its posterior responsibility. Let $\pi_i = (\pi_{i,1}, \dots, \pi_{i,K_{comp}})$ denote the mixture-proportion vector of slide i , where $\pi_{i,r}$ represents the fraction of patches assigned to component r among K_{comp} grouped components. To model structural co-occurrence patterns, a density $G_m^{comp,c}$ is fitted over the simplex-valued vectors $\{\pi_i\}$ for slides of class c . During pseudo-bag synthesis, sampled proportions $\tilde{\pi}$ from $G_m^{comp,c}$ control how many instances are drawn from each component, ensuring that synthetic bags preserve realistic intra-bag composition.

Top- K Discriminative Attribute Modeling. To emphasize highly discriminative regions, a Top- K GMM G_m^{top} is constructed. After an initial local MIL pretraining, the top- K' patches with the highest MIL attention scores are selected for each slide with class c . The top- K discriminative attribute of class- c slides in center m and modeled as:

$$p(z \mid \text{top-}K') = \sum_{r=1}^{K_{top}} \beta_r \mathcal{N}(z \mid \nu_r, \Lambda_r), \quad (4)$$

where $\{\beta_r, \nu_r, \Lambda_r\}$ are mixture weights, means and covariances. G_m^{top} captures class-specific discriminative patterns that are injected into pseudo-bags.

By integrating the above attributes, $\mathcal{G}_m$ provides a multi-attribute GMM group for center m . Instead of sharing raw data, each center uploads only $\mathcal{G}_m$ to the server, which then redistributes all $\{\mathcal{G}_{m'}\}_{m' \neq m}$ back to each center, enabling cross-center generative knowledge transfer under strict privacy constraints.

2.3 Cross-Domain Pseudo-Bag Generation

Leveraging the shared GMM groups $\{\mathcal{G}_{m'}\}$, each center synthesizes cross-domain pseudo-bags before federated training. Two complementary strategies are introduced: (i) distribution replay, which injects external morphological diversity by directly sampling from other GMMs, and (ii) covariance-aligned feature mapping, which transforms local patch features into the statistical space of another center. Replay enhances sample diversity, whereas mapping provides precise distribution alignment.

Distribution Replay via GMM Sampling. For center m , pseudo-bags of class c can be synthesized by sampling from the GMM group of another center $s \neq m$. A bag size $\tilde{n}$ is first drawn from the target center’s patch-count model G_s^{pc} , followed by a composition vector $\tilde{\pi}$ sampled from the class-specific composition model $G_s^{comp,c}$, where each entry $\tilde{\pi}_r$ determines the proportion of instances drawn from component r . Accordingly, $\tilde{n}_r = \lfloor \tilde{\pi}_r \tilde{n} \rfloor$ feature vectors are sampled from the corresponding class-wise GMM $G_s^{cls,c}$. Additional high-attention features from the Top- K GMM G_s^{top} can be incorporated to strengthen class-discriminative cues. Repeating this procedure produces the replay pseudo-bag set $\tilde{\mathcal{D}}_m^{(1)}$, which preserves class balance while reflecting the WSI statistics of external centers.

Covariance-Aligned Feature Mapping. Although distribution replay introduces cross-domain diversity, feature statistics across centers may still remain misaligned. To explicitly correct such discrepancies, each class-wise GMM $G_m^{cls,c}$ is paired with its counterpart $G_s^{cls,c}$ from a target center s . Component correspondence is determined by minimizing the 2-Wasserstein distance between Gaussian components, yielding a mapping $k \mapsto j(k)$ from local to target components. For a local patch feature z assigned to component k , its aligned representation is obtained via:

$$z' = A_{k \rightarrow j(k)} \left(z - \mu_{c,k}^{(m)} \right) + \mu_{c,j(k)}^{(s)}, \quad (5)$$

where $A_{k \rightarrow j(k)}$ is a whitening-coloring transform that matches the second-order statistics of the two components. Applying this mapping to all instances in a local bag produces a covariance-aligned pseudo-bag $\tilde{Z}^{(2)}$, and repeating the process yields the mapped pseudo-bag set $\tilde{\mathcal{D}}_m^{(2)}$, which complements replay by enforcing distribution-level alignment.

Replayed pseudo-bags $\tilde{\mathcal{D}}_m^{(1)}$ and mapped pseudo-bags $\tilde{\mathcal{D}}_m^{(2)}$ jointly augment the local dataset: $\hat{\mathcal{D}}_m = \mathcal{D}_m \cup \tilde{\mathcal{D}}_m^{(1)} \cup \tilde{\mathcal{D}}_m^{(2)}$.

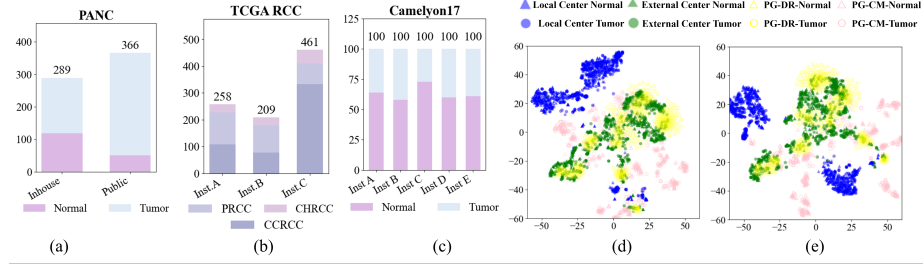


Fig. 2. (a–c) Sample distributions of the three multi-center datasets. (d–e) T-SNE visualization of patch-level features: (d) PANC and (e) CAMELYON17.

2.4 Federated Optimization with FedGMM-Bag

After constructing the augmented dataset $\hat{\mathcal{D}}_m$ that includes both real bags and pseudo-bags generated from external distributions. Federated optimization proceeds as follows. At the beginning of each round, the server broadcasts the current global MIL model to all centers. Each center performs local training on its augmented dataset for several epochs and returns updated parameters. The server aggregates the local updates using dataset-size weights, consistent with FedAvg[15], and redistributes the refined model parameters for the next round. The global objective optimized by FedGMM-Bag is defined as:

$$\min_{\theta} \mathcal{L}(\theta) = \sum_{m=1}^M \frac{N_m}{N_{\text{tot}}} \mathbb{E}_{(Z,y) \sim \hat{\mathcal{D}}_m} [\ell(f(Z; \theta), y)], \quad (6)$$

where $\ell(\cdot, \cdot)$ is the slide-level cross-entropy loss, N_m is the total number of bags at center m , including both real bags and generated pseudo-bags.

3 Experiment

3.1 Dataset and Implementation Details

Dataset. We evaluate our method on three multi-center histopathology WSI datasets, and their sample distributions are illustrated in Fig. 2 (a)–(c). **PANC** consists of pancreatic WSIs from our in-house cohort and the public TCGA-PANC [18] dataset, covering benign and malignant cases. **TCGA-RCC** [2] contains three renal cell carcinoma subtypes (CCRCC, PRCC, CHRC) and is partitioned into centers following the HistoFS [19] protocol. Both datasets use a 7:1.5:1.5 train/val/test split per center. **CAMELYON17** [1] provides lymph node WSIs from five centers, where A–C are used for training and D–E as used as unseen test centers in our experiments.

Implementation Details. All WSIs are preprocessed via CLAM [13] pipeline, tiled into 256×256 patches at $20\times$, and encoded using ImageNet pretrained

Table 1. Main results on PANC and RCC. Results are mean_{std} over five folds. SC and JT denote the lower and oracle upper bounds. Bold marks the best federated method.

Method	PANC			RCC			
	Inst.A	Inst.B	Entropy $\downarrow$	Inst.A	Inst.B	Inst.C	Entropy $\downarrow$
SC _{Lowerbound}	79.37 _{6.74}	80.32 _{7.93}	0.5185 _{0.04}	94.73 _{4.37}	93.92 _{3.91}	90.42 _{3.18}	0.2904 _{0.03}
JT _{Upperbound}	83.26 _{4.62}	88.58 _{5.96}	0.3750 _{0.03}	99.37 _{3.02}	98.87 _{2.39}	98.69 _{2.22}	0.0440 _{0.01}
DTFD-MIL [23]	80.37 _{6.29}	81.19 _{7.52}	0.4975 _{0.04}	95.27 _{4.25}	94.68 _{3.67}	92.18 _{3.37}	0.2496 _{0.03}
PseMix [11]	79.96 _{6.34}	80.92 _{6.95}	0.5051 _{0.04}	95.14 _{4.04}	94.23 _{3.54}	91.97 _{3.07}	0.2607 _{0.03}
HistoFL [12]	80.21 _{5.26}	81.29 _{5.43}	0.4981 _{0.04}	97.58 _{4.33}	95.03 _{2.95}	97.38 _{2.96}	0.1417 _{0.02}
MOON [9]	82.13 _{5.89}	83.58 _{6.37}	0.4495 _{0.04}	98.13 _{3.81}	96.06 _{3.08}	98.31 _{3.16}	0.1066 _{0.02}
DACS [20]	82.87 _{5.37}	85.86 _{6.13}	0.4135 _{0.03}	98.26 _{3.96}	96.28 _{3.12}	98.19 _{2.65}	0.1034 _{0.02}
FedWSIDD [7]	82.94 _{5.94}	86.75 _{6.29}	0.4017 _{0.03}	98.31 _{4.02}	96.31 _{2.52}	98.27 _{2.49}	0.1012 _{0.02}
HistoFS [19]	83.06 _{5.01}	86.93 _{5.72}	0.3981 _{0.03}	99.25 _{3.73}	96.39 _{2.48}	98.46 _{2.34}	0.0840 _{0.02}
Ours	83.11_{4.88}	86.95_{5.38}	0.3972_{0.03}	99.29_{3.31}	96.83_{2.43}	98.51_{2.30}	0.0765_{0.02}

Table 2. Results on the CAMELYON17 dataset. Models are trained on centers A–C and evaluated on unseen centers D–E. Best results are highlighted in bold.

Method	CAMELYON17			
	Inst. D	Inst E	Avg.	Entropy $\downarrow$
SC _{Lowerbound}	68.22 _{6.28}	72.37 _{8.04}	70.29 _{7.39}	0.7140 _{0.08}
JT _{Upperbound}	83.95 _{4.32}	85.28 _{5.18}	84.62 _{4.96}	0.4078 _{0.03}
DTFD-MIL [23]	73.39 _{5.75}	76.20 _{7.83}	74.86 _{6.83}	0.6264 _{0.06}
PseMix [11]	72.56 _{5.82}	75.82 _{7.66}	74.19 _{6.72}	0.6386 _{0.06}
HistoFL [12]	74.69 _{5.27}	78.93 _{6.95}	76.81 _{6.34}	0.5839 _{0.04}
MOON [9]	78.31 _{5.36}	80.02 _{6.83}	79.17 _{6.27}	0.5335 _{0.05}
DACS [20]	78.59 _{5.48}	81.34 _{6.49}	79.97 _{5.94}	0.5155 _{0.04}
FedWSIDD [7]	80.19 _{4.75}	82.02 _{5.82}	81.11 _{5.33}	0.4899 _{0.04}
HistoFS [19]	81.29 _{4.59}	83.42 _{5.96}	82.36 _{5.28}	0.4611 _{0.04}
Ours	82.18_{4.40}	83.87_{5.73}	83.03_{5.15}	0.4455_{0.04}

ResNet-50 [6] to obtain 1024-D features. CLAM-SB[13] is used as the local and global MIL backbone. For GMM groups, we use 4/8/4/4 components for $\mathcal{G}_m^{cp}$, $\mathcal{G}_m^{cls}$, $\mathcal{G}_m^{comp}$, and $\mathcal{G}_m^{topk}$, respectively, with diagonal covariances and EM optimization. Each center synthesizes the same number of pseudo-bags as its real WSIs, with an 80/20 split between distribution replay and covariance-aligned mapping. Federated training runs for 20 global rounds with 20 local epochs each, using Adam ($\text{lr}=1\text{e}-5$) and FedAvg [15] aggregation. All results are averaged over five seeds. **Communication Cost:** Each center transmits its GMM group only once, adding merely about 0.28 MB per center in our 1024-D, 3-class setting of TCGA-RCC (73,980 float32 parameters). The total overhead ($M \times 0.28\text{MB}$, $M = 3$) is negligible compared with standard federated updates. **Evaluation Metrics:** Performance is evaluated using the area under the ROC curve (AUC) across institutions. Entropy score H is also reported, which quantifies prediction uncertainty across centers and is defined as $H = -\sum_{m=1}^M \text{AUC}_m \cdot \log_e(\text{AUC}_m)$, a lower H indicates more consistent performance across institutions, making it a suitable metric for evaluating cross-site robustness.

Table 3. Ablation study on attribute-aware GMM components and pseudo-bag generation strategies.

Modules					PANC		RCC	
					AUC $\uparrow$	Entropy $\downarrow$	AUC $\uparrow$	Entropy $\downarrow$
					$\mathcal{G}_m^{cp}$	$\mathcal{G}_m^{cls}$	$\mathcal{G}_m^{comp}$	$\mathcal{G}_m^{topk}$
GMM Components					80.75	0.4981	96.66	0.1417
	✓	✓			82.46	0.4587	97.25	0.1171
	✓	✓	✓		83.17	0.4421	97.45	0.1087
	✓	✓	✓	✓	85.03	0.3972	98.21	0.0765
Pseudo-Bag	DR Only				84.43	0.4120	97.80	0.0941
Generation	CM Only				80.70	0.4993	94.08	0.2482

3.2 Comparison with State-of-the-Art Methods

Table 1 reports results on PANC and RCC. Single Center (SC) serves as a lower bound with purely local training, while Joint Training (JT) provides an impractical upper bound that requires centralized data sharing. Compared with pseudo-bag augmentation methods DTFD-MIL [23] and PseMix [11], federated approaches achieve higher AUC and lower entropy because local-only synthesis cannot model cross-center variability. General FL algorithms MOON [9] and DACS [20] improve over naive aggregation but remain suboptimal for pathology data. Pathology-oriented frameworks [12, 7, 19] perform strongly, yet FedGMM-Bag achieves the best overall performance. Table 2 summarizes results on unseen center in CAMELYON17. FedGMM-Bag consistently surpasses all FL and pseudo-bag baselines on both seen and unseen centers, demonstrating robustness under distribution shift and cross-domain generalization.

To qualitatively assess the effectiveness of the pseudo-bag generation strategies, we perform t-SNE visualization on patch features, as shown in Fig. 2(d)–(e). Real samples from local (blue) and external (green) centers exhibit clear domain shifts. Synthetic features generated by Distribution Replay (DR, yellow) successfully reproduce foreign distributions without accessing raw data, while Covariance-aligned Feature Mapping (CM, pink) transforms local samples toward the external style, effectively narrowing cross-domain discrepancies.

3.3 Ablation Study

We conduct ablation studies on both the attribute-aware GMM design and the pseudo-bag generation strategies. As shown in Table 3, using only patch-count and class-wise GMMs gives reasonable improvements compared with baseline HistoFL, while adding bag-composition and top- K GMMs progressively improves performance, highlighting the role of intra-bag structure and discriminative features. As for two pseudo-bag generation strategies, DR alone improves robustness and CM alleviates domain shifts by aligning feature statistics. Their combination achieves the highest AUC and lowest entropy.

4 Conclusion

This work presented FedGMM-Bag for federated learning of WSI classification. By modeling center-specific WSI attributes with multi-attribute GMMs and leveraging dual pseudo-bag generation strategies, our framework effectively addresses both privacy constraints and feature distribution heterogeneity across institutions. Extensive experiments on multiple multi-center datasets demonstrated consistent improvements in cross-domain generalization while maintaining low communication overhead.

Acknowledgments. This work is supported by National Natural Science Foundation of China (Grant No. 62475072, 42527804), Fundamental Research Funds for the Central Universities, the Science and Technology Commission of Shanghai Municipality (Grant No. 25xtcx00600, 22DZ2229004), China Postdoctoral Science Foundation (Grant No. 2025M783508).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bándi, P., Geessink, O., Manson, Q., Van Dijk, M., Balkenhol, M., Hermesen, M., Ehteshami Bejnordi, B., et al.: From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge. *IEEE Transactions on Medical Imaging* **38**(2), 550–560 (Feb 2019). <https://doi.org/10.1109/TMI.2018.2867350>
2. Cancer Genome Atlas Research Network, Linehan, W.M., Spellman, P.T., Rickerts, C.J., Creighton, C.J., Fei, S.S., Davis, C., Wheeler, D.A., Murray, B.A., et al.: Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *The New England Journal of Medicine* **374**(2), 135–145 (Jan 2016). <https://doi.org/10.1056/NEJMoa1505917>
3. Chen, H., Li, H., Zhang, Y., Bi, J., Zhang, G., Zhang, Y., Torr, P., Gu, J., Krompass, D., Tresp, V.: Fedbip: Heterogeneous one-shot federated learning with personalized latent diffusion models. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 30440–30450 (2025). <https://doi.org/10.1109/CVPR52734.2025.02834>
4. Fillioux, L., Boyd, J., Vakalopoulou, M., Cournède, P.H., Christodoulidis, S.: Structured State Space Models for Multiple Instance Learning in Digital Pathology (Jun 2023). <https://doi.org/10.48550/arXiv.2306.15789>
5. Guan, H., Yap, P.T., Bozoki, A., Liu, M.: Federated learning for medical image analysis: A survey. *Pattern recognition* **151**, 110424 (2024)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
7. Jin, H., Liu, S., Cong, C., Feng, Q., Liu, Y., Huang, L., Hu, Y.: Fedwsidd: Federated whole slide image classification via dataset distillation. *arXiv preprint arXiv:2506.15365* (2025)

8. Lee, M.: Recent advancements in deep learning using whole slide imaging for cancer prognosis. *Bioengineering* **10**(8) (2023)
9. Li, Q., He, B., Song, D.: Model-contrastive federated learning. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 10713–10722 (2021)
10. Liu, A., Li, T., Cai, J., Veer Vajrала, S.S.: FuzzyMIL: Decoupling Pathological Phenotypes through Deep Fuzzy Clustering for Efficient Whole Slide Image Analysis. In: *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp. 1–5 (Apr 2025). <https://doi.org/10.1109/ICASSP49660.2025.10889453>
11. Liu, P., Ji, L., Zhang, X., Ye, F.: Pseudo-Bag Mixup Augmentation for Multiple Instance Learning-Based Whole Slide Image Classification. *IEEE Transactions on Medical Imaging* **43**(5), 1841–1852 (May 2024). <https://doi.org/10.1109/TMI.2024.3351213>
12. Lu, M.Y., Chen, R.J., Kong, D., Lipkova, J., Singh, R., Williamson, D.F.K., Chen, T.Y., Mahmood, F.: Federated learning for computational pathology on gigapixel whole slide images. *Medical Image Analysis* **76**, 102298 (Feb 2022). <https://doi.org/10.1016/j.media.2021.102298>
13. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering* **5**(6), 555–570 (2021)
14. McGenity, C., Clarke, E.L., Jennings, C., Matthews, et al.: Artificial intelligence in digital pathology: A systematic review and meta-analysis of diagnostic test accuracy. *npj Digital Medicine* **7**(1), 114 (May 2024). <https://doi.org/10.1038/s41746-024-01106-8>
15. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: *Artificial intelligence and statistics*. pp. 1273–1282. PMLR (2017)
16. Peng, Z., Song, Y., Wang, Q., Xiao, X., Tang, Z.: Fedbn: A communication-efficient federated learning strategy based on blockchain. In: *2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. pp. 754–759. IEEE (2024)
17. Pettersson, S.Z., Liang, K.Y., Andresen, J.C.: Federated Gaussian Mixture Models (Jun 2025). <https://doi.org/10.48550/arXiv.2506.01780>
18. Raphael, B.J., Hruban, R.H., Aguirre, A.J., Moffitt, R.A., Yeh, J.J., Stewart, C., Robertson, A.G., Cherniack, A.D., Gupta, M., Getz, G., Gabriel, S.B., Meyerson, M., Cibulskis, C., Fei, S.S., et al.: Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **32**(2), 185–203.e13 (Aug 2017). <https://doi.org/10.1016/j.ccell.2017.07.007>
19. Raswa, F.H., Lu, C.S., Wang, J.C.: Histofs: Non-iid histopathologic whole slide image classification via federated style transfer with roi-preserving. *CVPR* (2025)
20. Yang, F., Zhong, Z., Luo, Z., He, Y., Li, S., Sebe, N.: Diversity-authenticity co-constrained stylization for federated domain generalization in person re-identification. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 38, pp. 6477–6485 (2024)
21. Yang, M.S., Lai, C.Y., Lin, C.Y.: A robust em clustering algorithm for gaussian mixture models. *Pattern Recognition* **45**(11), 3950–3961 (2012)
22. Yuan, X., Li, P.: On convergence of fedprox: Local dissimilarity invariant bounds, non-smoothness and beyond. *Advances in Neural Information Processing Systems* **35**, 10752–10765 (2022)

23. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: DTFD-MIL: Double-Tier Feature Distillation Multiple Instance Learning for Histopathology Whole Slide Image Classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18802–18812 (2022)
24. Zheng, S., Wen, S., Li, C., Chen, Q., Li, L.: Federated Hybrid-Supervised Learning for Universal Medical Image Segmentation. In: ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5 (Apr 2025). <https://doi.org/10.1109/ICASSP49660.2025.10887826>