FedWSIDD: Federated Whole Slide Image Classification via Dataset Distillation

Haolong Jin[1], Shenglin Liu[1⋆], Cong Cong[2], Qingmin Feng[1], Yongzhi Liu[1], Lina Huang[1], and Yingzi Hu[1]

[1] Union Hospital, Tongji Medical College, Huazhong University of Science and Technology

[2] Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2113, Australia

jhl_bme@hust.edu.cn

Abstract. Federated learning (FL) has emerged as a promising approach for collaborative medical image analysis, enabling multiple institutions to build robust predictive models while preserving sensitive patient data. In the context of Whole Slide Image (WSI) classification, FL faces significant challenges, including heterogeneous computational resources across participating medical institutes and privacy concerns. To address these challenges, we propose FedWSIDD, a novel FL paradigm that leverages dataset distillation (DD) to learn and transmit synthetic slides. On the server side, FedWSIDD aggregates synthetic slides from participating centres and distributes them across all centres. On the client side, we introduce a novel DD algorithm tailored to histopathology datasets which incorporates stain normalisation into the distillation process to generate a compact set of highly informative synthetic slides. These synthetic slides, rather than model parameters, are transmitted to the server. After communication, the received synthetic slides are combined with original slides for local tasks. Extensive experiments on multiple WSI classification tasks, including CAMELYON16 and CAMELYON17, demonstrate that FedWSIDD offers flexibility for heterogeneous local models, enhances local WSI classification performance, and preserves patient privacy. This makes it a highly effective solution for complex WSI classification tasks. The code is available at FedWSIDD.

Keywords: Federated Learning · Dataset Distillation · WSI classification.

1 Introduction

Computational pathology aims to enhance the precision of histopathological tissue examination and it is crucial for accurate disease diagnosis. Deep learning has demonstrated significant success in various digital pathology tasks; however, Whole Slide Images (WSIs) pose unique challenges. Their enormous size makes

⋆ Corresponding author: liushenglin@hust.edu.cn

direct model training impractical, and the labour-intensive nature of annotation often limits available labels to the slide level. To address these issues, weakly-supervised Multiple Instance Learning (MIL) frameworks [4,7] are commonly used, leveraging slide-level ground truth labels while treating each WSI as a collection of patches.

As with many machine learning applications, incorporating diverse data that reflect variations in patient populations and data collection protocols can significantly enhance MIL performance [20]. However, sharing medical data poses challenges, including regulatory and legal constraints, as well as technical obstacles like the high costs of data transfer and storage. Federated learning (FL) provides feasible solutions to the above issues by enabling algorithms to be trained on decentralised data across multiple institutions, allowing each institution to retain its data securely. FL-based WSI classification methods have been actively developed in recent years [15]. Though effective, they focus on patch-level classification which requires precise patch-level annotations. However, such annotations are usually hard to obtain in real clinical scenarios. Thus, FedHisto [20] was proposed to tackle the more practical slide-level classification task where the MIL classifiers are shared between the centres and server. Huang *et al.* [15] further enhanced the local model performance by introducing a dynamic parameter updating mechanism. However, these methods assume that centres have similar computational resources and identical MIL architectures, limiting their application in settings with heterogeneous MIL models.

Accordingly, we propose FedWSIDD, a novel federated learning (FL) algorithm for weakly supervised WSI classification that incorporates dataset distillation. To ensure compatibility across different MIL models, rather than employing model augmentation [30] or maintaining a diverse MIL model pool [5], we exploit the fact that pre-extracted features are MIL-agnostic. Hence, we explicitly design a dataset distillation [28,10] algorithm that matches the feature space between real and synthetic patches, enabling a more adaptable and efficient dataset distillation process. Furthermore, to mitigate the training challenges caused by the inherent heterogeneity in the appearance of H&E-stained tissue slides, we integrate stain normalisation [9,8,22] into the dataset distillation process. At each participating medical centre, dataset distillation is performed to generate synthetic slides, which are then transmitted to the central server. The server aggregates and redistributes these synthetic slides to all centres. Each centre subsequently incorporates the updated synthetic slides from other centres as local data augmentation in the next training round. Our contributions can be summarised as follows:

- We introduce FedWSIDD, a novel federated learning paradigm for weakly supervised WSI classification, where each centre learns and shares synthetic WSIs, enabling flexible MIL selection based on local computational resources.
- We design a dataset distillation algorithm tailored for WSI datasets, which integrates stain normalisation into the learning process of synthetic patches, enhancing the quality of the synthetic slides and leading to enhanced downstream classification performance.

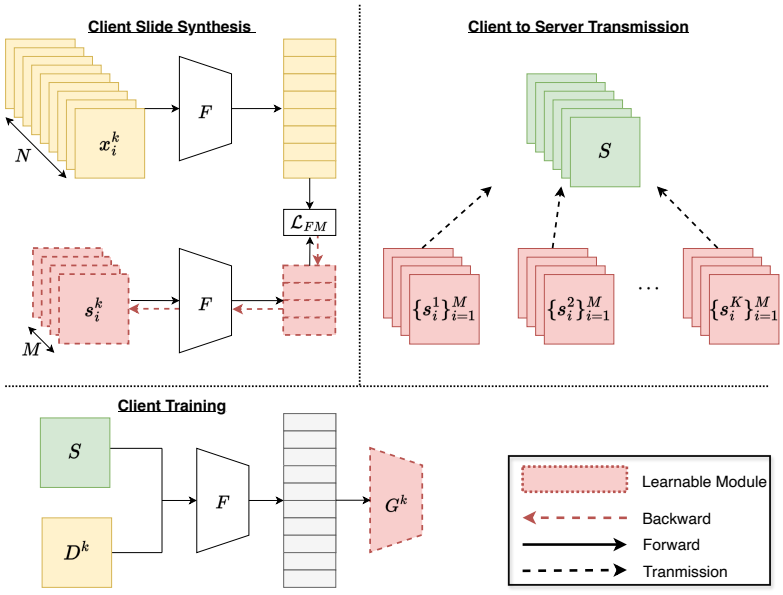

Fig. 1. FedWSIDD pipeline. Each client first generates a set of synthetic slides which are then transmitted to the server. The server gathers synthetic slides from all participating clients and distributes the aggregated synthetic slides to them. Upon receiving the synthetic slides, clients merge them with their original real slides for local training.

– Extensive experiments demonstrate the effectiveness of FedWSIDD. Compared to state-of-the-art FL methods, FedWSIDD achieves superior performance across various settings, including both homogeneous and heterogeneous local models.

2 Method

2.1 Preliminary

Given a WSI dataset consisting of N WSI slides $\{x_i, y_i\}_i^N$, we first crop x_i into a set of patches $\{c_t^i\}_{t=1}^T$, where T could vary across WSIs. Then a pretrained feature extractor ($\mathcal{F}$) is used to convert c_t^i to patch embedding $e_t^i = \mathcal{F}(c_t^i)$. This is usually done in an offline manner where e_t^i is then aggregated to form a bag representation that is handled by the MIL classifier (G) to obtain the slide-level prediction.

The standard Federated WSI classification involves multiple medical centres working collaboratively to enhance the performance of a global MIL model. In practical scenarios, different medical centres may have different computational resources and wish to adopt different feature extractors or MIL classifiers. Thus, we focus on the personalised federated WSI classification to tackle this heterogeneous local model scenario. Supposing we have K centres where each consists

of their own WSI datasets ($\{D^1, D^2, \cdots, D^K\}$) and the task can be formulated as:

$$\arg\min_{\{G^k\}_{k=1}^K} \mathcal{L} = \sum_{k=1}^{K} \frac{|D^k|}{|D|} \mathbb{E}_{(x,y)\in D^k} \mathcal{L}(G^k; x, y) \tag{1}$$

here D represents the combined dataset from all centres, G^k denotes the MIL classifier selected by centre k and $\mathcal{L}$ is a task-specific loss function, and different MIL methods define $\mathcal{L}$ in various forms. For example, in CLAM [21], $\mathcal{L}$ comprises both slide-level and patch-level loss components. If all centres choose to use the same MIL classifier, this scenario becomes a homogeneous local model setup and standard FL methods can be applied.

2.2 Federated WSI Learning with Synthetic Slides

We illustrate the pipeline of FedWSIDD in Fig. 1. Since each centre may have different computational resources and employ a distinct MIL classifier G^k, transmitting model weights or gradients, as done in traditional FL methods [20], is neither practical nor efficient. To overcome this limitation, we propose generating a set of synthetic slides and transmitting $\{s_i\}_{i=1}^M$, where M represents the number of synthetic slides per class and satisfies $M \ll N$, instead of exchanging model weights.

Specifically, each participating centre k first performs local synthetic slide generation using dataset distillation (see Sec. 2.3 for details) to produce M synthetic slides, denoted as $\{s_i^k\}_{i=1}^M$. These synthetic slides are then uploaded to the server, where the global synthetic slides (S) are aggregated as:

$$S = \bigcup_{k\in K} \{s_i^k\}_{i=1}^M \tag{2}$$

The global synthetic slides S are subsequently distributed back to the participating centres, where they are merged with real local slides for local MIL training.

2.3 Local Synthetic Slides with Dataset Distillation

Generating synthetic WSIs directly from real slides is highly challenging due to their gigapixel resolution. To address this, we represent each synthetic slide s_i as a collection of synthetic patches, *i.e.,* $s_i = \{p_j\}_{j=1}^B$, where $p_j \in \mathbb{R}^{C\times H\times W}$ and $B \ll T$, with each patch treated as a set of learnable parameters. To ensure compatibility across different MIL models, we leverage the fact that pre-extracted features are ML-agnostic. Thus, we explicitly match the feature space between real and synthetic patches, enabling a more adaptable and efficient dataset distillation process. Specifically, during each distillation iteration, we iterate through all classes. For each class, we randomly sample one real slide (x_j) and one synthetic slide (s_j). To mitigate training difficulties caused by the heterogeneity of H&E-stained tissue slides, we incorporate stain normalisation [22] into the training process. Specifically, a template patch with pixel mean and standard deviation closest to the dataset average is selected, and its stain vector and

concentration are pre-computed offline. During training, stain normalisation is applied via matrix multiplication, aligning both real patches $\{c_t^j\}_{t=1}^T$ and the synthetic patches $\{p_b^j\}_{b=1}^B$ using the derived stain vector and concentration and perform feature matching using:

$$\arg\min_{\{p_b^j\}_{b=1}^B} \mathcal{L}_{FM} = \left[\left\| \frac{1}{T} \sum_{i=0}^{T} \mathcal{F}(c_t^j) - \frac{1}{B} \sum_{b=0}^{B} \mathcal{F}(p_b^j) \right\|^2 \right], \tag{3}$$

Note that, unlike previous works [14] that apply dataset distillation to federated learning, we are the first to successfully extend this concept to WSI classification, providing a potential solution for real-world clinical applications. Furthermore, we incorporate stain normalisation into the learning process of synthetic patches, enhancing dataset distillation algorithms specifically for histopathology images. Empirically, we demonstrate that FedWSIDD outperforms these existing approaches on WSI datasets.

3 Experiments

3.1 Datasets

CAMELYON16 [3] is for detecting breast cancer metastasis in lymph node sections. The task is binary classification, distinguishing between normal and tumour slides. The dataset includes 399 WSIs from two medical centres: RUMC (C1), with 169 slides for training (99 normal/70 tumour) and 74 slides for testing (50 normal/24 tumour); and UMCU (C2), contributing 101 slides for training (60 normal/41 tumour) and 55 slides for testing (31 normal/24 tumour).
CAMELYON17 [2] consists WSIs collected from five medical centres. As test labels are not officially provided, we used the training set (500 slides from 100 patients) for our experiments. The task is to classify each slide into one of four categories: negative (neg), isolated tumour cells (itc), micro-metastases (micro), and macro-metastases (macro). Each centre (C1–C5) includes 20 patients, with five slides per patient. The class distribution across centres is as follows: C1: 11 itc, 10 micro, 15 macro, and 64 neg; C2: 7 itc, 23 micro, 12 macro, and 58 neg; C3: 2 itc, 8 micro, 15 macro, and 75 neg; C4: 8 itc, 5 micro, 26 macro, and 61 neg; C5: 8 itc, 26 micro, 5 macro, and 61 neg. Within each centre, We randomly allocated 80% of data for training and 20% for testing, ensuring that slides from the same patient were assigned to the same split.

3.2 Implementation Details

We loaded WSIs at 20x for CAMELYON17, and 40x for CAMELYON16. Moreover, we extracted 256x256 patches from CAMELYON16 and CAMELYON17. For patch feature extraction, we employed a ResNet50 [12] which is pre-trained on ImageNet [11]. Each patch was resized to 224×224 pixels for feature extraction. For the FL setup, we follow [14,25] and conduct a one-shot communication,

Table 1. Homogeneous model performance comparison between FedWSIDD and *standard FL* methods on CAMELYON16 (**CA16**), CAMELYON17 (**CA17**).

Methods		Local	FedHisto	FedDyn	MOON	FedImpro	FedMut	FedWSIDD
CA16	C1	$85.2_{\pm 1.0}$	$92.8_{\pm 1.7}$	$92.3_{\pm 0.6}$	$91.0_{\pm 1.7}$	$92.1_{\pm 1.4}$	$93.6_{\pm 0.5}$	$\mathbf{94.2}_{\pm 0.2}$
	C2	$74.0_{\pm 1.6}$	$74.8_{\pm 1.8}$	$84.3_{\pm 1.8}$	$78.0_{\pm 1.7}$	$80.3_{\pm 0.8}$	$82.4_{\pm 1.2}$	$84.5_{\pm 0.1}$
	Avg	$79.6_{\pm 0.6}$	$83.8_{\pm 1.6}$	$88.3_{\pm 0.7}$	$84.6_{\pm 0.3}$	$86.5_{\pm 0.8}$	$88.6_{\pm 0.6}$	$\mathbf{90.1}_{\pm 0.2}$
	p-value	0.00001	0.0001	0.002	0.0002	0.0008	0.004	-
CA17	C1	$80.0_{\pm 0.0}$	$85.0_{\pm 4.1}$	$80.0_{\pm 0.0}$	$\mathbf{90.0}_{\pm 0.0}$	$82.8_{\pm 1.5}$	$82.0_{\pm 2.0}$	$86.0_{\pm 2.0}$
	C2	$60.0_{\pm 4.1}$	$71.7_{\pm 2.4}$	$71.7_{\pm 2.4}$	$61.7_{\pm 2.4}$	$70.2_{\pm 4.8}$	$73.0_{\pm 2.4}$	$\mathbf{76.0}_{\pm 3.7}$
	C3	$86.7_{\pm 2.4}$	$81.7_{\pm 2.4}$	$83.3_{\pm 4.7}$	$83.3_{\pm 2.4}$	$88.8_{\pm 0.9}$	$88.0_{\pm 2.4}$	$\mathbf{89.0}_{\pm 2.0}$
	C4	$60.0_{\pm 0.0}$	$61.7_{\pm 2.4}$	$58.3_{\pm 2.4}$	$63.3_{\pm 2.4}$	$66.2_{\pm 2.9}$	$70.6_{\pm 3.4}$	$\mathbf{71.0}_{\pm 2.0}$
	C5	$80.0_{\pm 0.0}$	$81.7_{\pm 4.7}$	$78.3_{\pm 2.4}$	$81.7_{\pm 2.4}$	$83.0_{\pm 1.4}$	$80.6_{\pm 3.1}$	$\mathbf{84.0}_{\pm 2.0}$
	Avg	$73.3_{\pm 0.9}$	$76.3_{\pm 0.5}$	$73.3_{\pm 0.9}$	$78.3_{\pm 1.3}$	$79.3_{\pm 1.0}$	$78.7_{\pm 2.4}$	$\mathbf{81.2}_{\pm 1.2}$
	p-value	0.0005	0.002	0.0003	0.02	0.03	0.04	-

Table 2. Homogeneous model performance comparison between FedDFP and *personalised FL* methods on CAMELYON16 (**CA16**), CAMELYON17 (**CA17**).

Methods		FedDM	FedProto	DESA	FedD3	FedDGM	SGPT	FedWSIDD
CA16	C1	$91.9_{\pm 0.8}$	$92.8_{\pm 0.6}$	$92.3_{\pm 0.6}$	$93.2_{\pm 1.3}$	$92.9_{\pm 1.1}$	$90.1_{\pm 0.9}$	$\mathbf{94.2}_{\pm 0.2}$
	C2	$81.3_{\pm 0.6}$	$83.6_{\pm 0.9}$	$83.6_{\pm 0.9}$	$82.3_{\pm 1.1}$	$83.8_{\pm 1.1}$	$83.9_{\pm 0.8}$	$\mathbf{84.5}_{\pm 0.1}$
	Avg	$86.6_{\pm 0.7}$	$88.2_{\pm 0.7}$	$88.0_{\pm 0.8}$	$87.8_{\pm 1.2}$	$88.4_{\pm 0.9}$	$87.0_{\pm 0.9}$	$\mathbf{90.1}_{\pm 0.2}$
	p-value	0.002	0.02	0.009	0.006	0.02	0.002	-
CA17	C1	$82.0_{\pm 2.4}$	$88.3_{\pm 2.4}$	$\mathbf{90.0}_{\pm 0.0}$	$83.0_{\pm 2.4}$	$83.0_{\pm 2.5}$	$75.0_{\pm 0.0}$	$86.0_{\pm 2.0}$
	C2	$67.0_{\pm 2.4}$	$75.0_{\pm 4.1}$	$75.0_{\pm 0.0}$	$67.0_{\pm 6.0}$	$73.0_{\pm 2.7}$	$68.3_{\pm 2.4}$	$\mathbf{76.0}_{\pm 3.7}$
	C3	$82.0_{\pm 2.5}$	$85.0_{\pm 0.0}$	$76.6_{\pm 4.7}$	$90.0_{\pm 0.0}$	$84.0_{\pm 2.2}$	$88.3_{\pm 2.4}$	$\mathbf{90.0}_{\pm 0.0}$
	C4	$64.0_{\pm 3.7}$	$65.0_{\pm 0.0}$	$61.7_{\pm 2.4}$	$65.0_{\pm 3.0}$	$71.0_{\pm 2.0}$	$\mathbf{72.0}_{\pm 4.1}$	$71.0_{\pm 2.0}$
	C5	$85.0_{\pm 0.0}$	$80.0_{\pm 0.0}$	$78.3_{\pm 2.4}$	$81.0_{\pm 2.0}$	$84.0_{\pm 2.0}$	$75.0_{\pm 2.4}$	$\mathbf{84.0}_{\pm 2.0}$
	Avg	$76.0_{\pm 1.6}$	$78.7_{\pm 1.1}$	$76.3_{\pm 1.5}$	$77.2_{\pm 1.0}$	$78.9_{\pm 1.0}$	$76.7_{\pm 2.6}$	$\mathbf{81.2}_{\pm 0.8}$
	p-value	0.02	0.04	0.01	0.001	0.04	0.02	-

1000 distillation rounds and 50 local MIL training rounds. Synthetic slides and MIL models were trained using the Adam optimiser with a learning rate of 0.0003, on a single Nvidia V100 GPU. The synthetic slides were randomly initialised and we set $M = 10$ and $B = 100$ with a size of 64x64. For evaluation, we measured test-set accuracy for each centre and the globally averaged accuracy, following the setup in [26,14]. Each experiment was repeated five times with fixed dataset splits to evaluate the impact of random initialisation and training variations, reporting the mean and standard deviation of classification accuracy. To further validate the improvement's significance, we conducted a paired t-test under the null hypothesis that FedWSIDD exhibits no significant performance difference compared to the baselines.

3.3 Comparison with State-of-the-arts

Baselines We compare FedWSIDD with two categories of federated learning (FL) methods. *Standard FL* involves training a global model through iterative local training and global aggregation. Specifically, we choose FedHisto [20], an enhanced version of FedAvg [23], FedDyn [1], MOON [19], FedImpro [27] and FedMut [13]. In the *Personalised FL* (pFL) category, which enables centres to develop customised models, we choose FedDM [29], FedProto [26], FedHE [6], DESA [14], FedD3 [25], FedDGM [17] and SGPT [18]. Additionally, we include a setup where each client trains solely on its own local data without FL (Local).

Table 3. Heterogeneous model performance comparison between FedDFP and *personalised FL* methods on CAMELYON16 (**CA16**), CAMELYON17 (**CA17**).

Methods		FedDM	FedHE	DESA	FedD3	FedDGM	SGPT	FedWSIDD
CA16	C1	$78.3_{\pm 0.9}$	$75.2_{\pm 2.2}$	$80.8_{\pm 2.6}$	$80.9_{\pm 0.8}$	$82.5_{\pm 0.9}$	$80.6_{\pm 1.2}$	$\mathbf{86.0}_{\pm 0.6}$
	C2	$64.4_{\pm 1.1}$	$66.1_{\pm 3.9}$	$67.6_{\pm 3.4}$	$70.9_{\pm 0.9}$	$73.0_{\pm 0.8}$	$71.1_{\pm 1.8}$	$\mathbf{83.7}_{\pm 0.9}$
	Avg	$71.4_{\pm 0.9}$	$70.2_{\pm 2.0}$	$75.0_{\pm 2.7}$	$76.0_{\pm 1.1}$	$77.6_{\pm 1.1}$	$76.2_{\pm 1.6}$	$\mathbf{84.8}_{\pm 0.7}$
	p-value	0.00001	0.00004	0.002	0.002	0.006	0.003	-
CA17	C1	$72.0_{\pm 6.2}$	$75.0_{\pm 0.0}$	$72.0_{\pm 6.2}$	$69.0_{\pm 4.2}$	$74.0_{\pm 3.2}$	$75.0_{\pm 0.0}$	$\mathbf{77.0}_{\pm 2.1}$
	C2	$73.0_{\pm 2.1}$	$61.7_{\pm 2.3}$	$70.0_{\pm 4.1}$	$73.0_{\pm 2.1}$	$72.0_{\pm 2.1}$	$68.3_{\pm 2.4}$	$\mathbf{73.0}_{\pm 2.1}$
	C3	$84.0_{\pm 3.7}$	$80.0_{\pm 0.0}$	$\mathbf{92.0}_{\pm 2.8}$	$84.0_{\pm 4.2}$	$85.0_{\pm 4.5}$	$88.3_{\pm 2.4}$	$86.0_{\pm 2.1}$
	C4	$65.0_{\pm 3.7}$	$73.0_{\pm 2.4}$	$67.0_{\pm 2.4}$	$69.0_{\pm 4.2}$	$67.0_{\pm 2.1}$	$71.6_{\pm 6.2}$	$\mathbf{73.0}_{\pm 2.1}$
	C5	$74.0_{\pm 3.7}$	$75.0_{\pm 4.1}$	$75.0_{\pm 4.1}$	$72.0_{\pm 2.1}$	$72.0_{\pm 2.1}$	$75.0_{\pm 4.1}$	$\mathbf{79.0}_{\pm 2.1}$
	Avg	$73.6_{\pm 3.6}$	$73.0_{\pm 0.0}$	$75.7_{\pm 2.1}$	$73.2_{\pm 5.8}$	$74.2_{\pm 4.1}$	$76.7_{\pm 2.0}$	$\mathbf{77.6}_{\pm 1.4}$
	p-value	0.03	0.01	0.03	0.01	0.04	0.04	-

Homogeneous Local Model This setup assumes that each medical centre has a similar computational capacity and is willing to adopt a unified model for their local tasks. Results are shown in Table. 1 and Table. 2, where we use one of the most commonly used MIL classifiers in the field, CLAM [21], as the global model. All FL methods outperform Local, highlighting that inter-centre collaboration can enhance histopathological diagnosis. Compared to standard FL methods, our FedWSIDD achieves the highest global average accuracy. Compared to personalised FLs, FedWSIDD shows superior performance, outperforming the previous works that adopt dataset distillation for federated learning [29,14,25,17]. This demonstrates the effectiveness of FedWSIDD for federated WSI classification.

Heterogeneous Local Model This setup assumes that different medical centres have varying computational resources, leading them to deploy different MIL classifiers. To simulate this scenario, we define a MIL model pool comprising CLAM [21], TransMIL [24], and ABMIL [16]. Additionally, since class embedding dimensions vary across different MIL classifiers, FedProto cannot be applied. Instead, we use FedHE [6], which shares class logits, for comparison. Each centre is randomly assigned a classifier from this pool. As shown in Table 3, FedWSIDD not only outperforms pFL methods but also demonstrates greater stability with a lower standard deviation. Furthermore, the p-values for all comparisons are below 0.05, confirming that FedWSIDD's improvements are statistically significant.

3.4 Ablation studies

Evaluation of synthetic slide dimension The dimension of the synthetic slide is a crucial hyperparameter, as it not only determines the number of parameters transmitted across clients but also affects the learning complexity of FedWSIDD. We present the performance of FedWSIDD on two datasets under different combinations of the number of synthetic slides per class (M) and the number of synthetic patches per slide (B) in Fig. 2. The results suggest that blindly increasing M and B does not necessarily lead to improved performance. We attribute this to the limited data variation in WSI datasets, where a smaller set of synthetic samples may suffice to capture the discriminative features of the

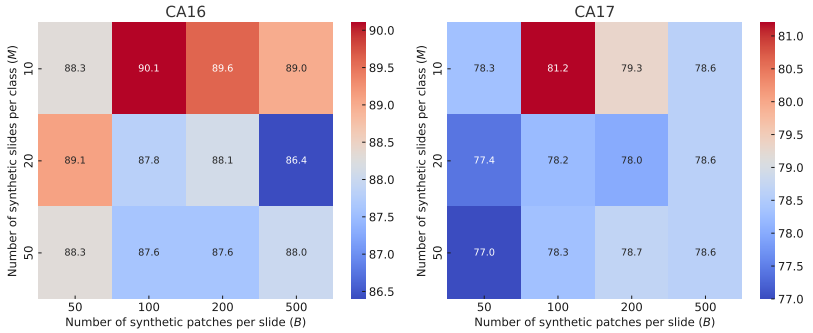

Fig. 2. Global averaged accuracy (%) across centres in CAMELYON16 (CA16) and CAMELYON17 (CA17) with different M and B.

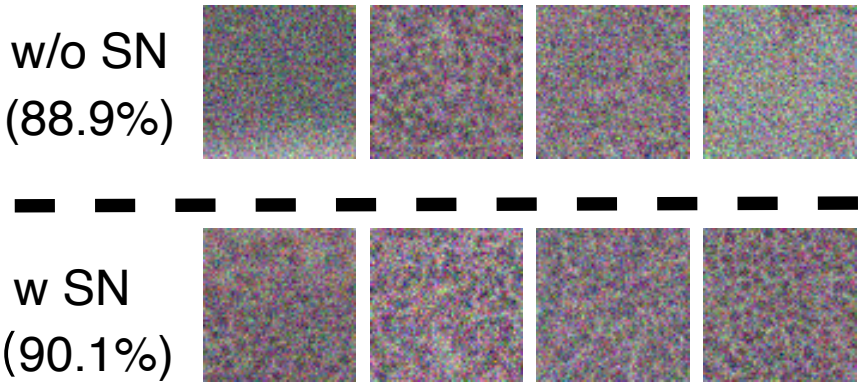

Fig. 3. Performance comparison with and without using stain normalisation during dataset distillation.

entire dataset. Furthermore, increasing M and B imposes higher computational demands and may hinder the convergence of FedWSIDD.

Effectiveness of stain normalisation We compare the global average accuracy across centres on CAMELYON16 with and without stain normalisation in Fig. 3. The results demonstrate that stain normalisation consistently improves the quality of synthetic samples, leading to enhanced model performance. This underscores its crucial role in making dataset distillation algorithms more adaptable to whole-slide image (WSI) datasets. Moreover, since the distilled samples are synthetic images, they offer the advantage of enhancing privacy by desensitising sensitive information. However, this synthetic nature may also reduce their explainability to some extent.

4 Conclusion

In conclusion, FedWSIDD offers a novel personalised federated learning framework for WSI classification, effectively addressing the challenges of computational constraints in medical centres. By exchanging synthetic slides, FedWSIDD achieves state-of-the-art performance with rapid convergence. Extensive results across multiple WSI datasets show that our approach not only enhances accuracy but also adapts well to both homogeneous and heterogeneous model settings, making it a promising solution for federated learning in resource-limited digital pathology applications.

Acknowledgments. The authors acknowledge supports from the Key Research and Development Plan of Hubei Province (2022BCA025). This work was partially supported by an NHMRC Ideas Grant (GNT202035).

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

References

1. Acar, D.A.E., Zhao, Y., Navarro, R.M., Mattina, M., Whatmough, P.N., Saligrama, V.: Federated learning based on dynamic regularization. arXiv preprint arXiv:2111.04263 (2021)
2. Bandi, P., Geessink, O., Manson, Q., Van Dijk, M., Balkenhol, M., Hermsen, M., Bejnordi, B.E., Lee, B., Paeng, K., Zhong, A., 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 (2019)
3. Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermsen, M., Manson, Q.F., Balkenhol, M., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA **318**(22), 2199–2210 (2017)
4. Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature medicine **25**(8), 1301–1309 (2019)
5. Cazenavette, G., Wang, T., Torralba, A., Efros, A.A., Zhu, J.Y.: Dataset distillation by matching training trajectories. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4750–4759 (2022)
6. Chan, Y.H., Ngai, E.C.: Fedhe: Heterogeneous models and communication-efficient federated learning. In: 2021 17th International Conference on Mobility, Sensing and Networking (MSN). pp. 207–214. IEEE (2021)
7. Chen, R.J., Chen, C., Li, Y., Chen, T.Y., Trister, A.D., Krishnan, R.G., Mahmood, F.: Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16144–16155 (2022)

8. Cong, C., Liu, S., Di Ieva, A., Pagnucco, M., Berkovsky, S., Song, Y.: Texture enhanced generative adversarial network for stain normalisation in histopathology images. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 1949–1952. IEEE (2021)
9. Cong, C., Liu, S., Di Ieva, A., Pagnucco, M., Berkovsky, S., Song, Y.: Colour adaptive generative networks for stain normalisation of histopathology images. Medical Image Analysis **82**, 102580 (2022)
10. Cong, C., Xuan, S., Liu, S., Pagnucco, M., Zhang, S., Song, Y.: Dataset distillation for histopathology image classification. arXiv preprint arXiv:2408.09709 (2024)
11. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
12. He, K., Zhang, X., et al.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)
13. Hu, M., Cao, Y., Li, A., Li, Z., Liu, C., Li, T., Chen, M., Liu, Y.: Fedmut: Generalized federated learning via stochastic mutation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 12528–12537 (2024)
14. Huang, C.Y., Srinivas, K., Zhang, X., Li, X.: Overcoming data and model heterogeneities in decentralized federated learning via synthetic anchors. arXiv preprint arXiv:2405.11525 (2024)
15. Huang, L., Shao, L., Bao, M., Guo, C., Shao, Z., Huang, X., Wang, M., Jiang, X., Hu, S.: Federated learning with comparative learning-based dynamic parameter updating on glioma whole slide images. Engineering Applications of Artificial Intelligence **137**, 109233 (2024)
16. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
17. Jia, Y., Vahidian, S., Sun, J., Zhang, J., Kungurtsev, V., Gong, N.Z., Chen, Y.: Unlocking the potential of federated learning: The symphony of dataset distillation via deep generative latents. In: European Conference on Computer Vision. pp. 18–33. Springer (2024)
18. Jia, Y., Vahidian, S., Sun, J., Zhang, J., Kungurtsev, V., Gong, N.Z., Chen, Y.: Unlocking the potential of federated learning: The symphony of dataset distillation via deep generative latents. In: European Conference on Computer Vision. pp. 18–33. Springer (2025)
19. 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)
20. Lu, M.Y., Chen, R.J., Kong, D., Lipkova, J., Singh, R., Williamson, D.F., Chen, T.Y., Mahmood, F.: Federated learning for computational pathology on gigapixel whole slide images. Medical image analysis **76**, 102298 (2022)
21. 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)
22. Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E.: A method for normalizing histology slides for quantitative analysis. In: IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1107–1110 (2009)
23. 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)

24. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Advances in neural information processing systems **34**, 2136–2147 (2021)
25. Song, R., Liu, D., Chen, D.Z., Festag, A., Trinitis, C., Schulz, M., Knoll, A.: Federated learning via decentralized dataset distillation in resource-constrained edge environments. In: 2023 International Joint Conference on Neural Networks (IJCNN). pp. 1–10. IEEE (2023)
26. Tan, Y., Long, G., Liu, L., Zhou, T., Lu, Q., Jiang, J., Zhang, C.: Fedproto: Federated prototype learning across heterogeneous clients. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 8432–8440 (2022)
27. Tang, Z., Zhang, Y., Shi, S., Tian, X., Liu, T., Han, B., Chu, X.: Fedimpro: Measuring and improving client update in federated learning. arXiv preprint arXiv:2402.07011 (2024)
28. Wang, T., Zhu, J.Y., Torralba, A., Efros, A.A.: Dataset distillation. arXiv preprint arXiv:1811.10959 (2018)
29. Xiong, Y., Wang, R., Cheng, M., Yu, F., Hsieh, C.J.: Feddm: Iterative distribution matching for communication-efficient federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16323–16332 (2023)
30. Zhang, L., Zhang, J., Lei, B., Mukherjee, S., Pan, X., Zhao, B., Ding, C., Li, Y., Xu, D.: Accelerating dataset distillation via model augmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11950–11959 (2023)