

FedSDA: Federated Stain Distribution Alignment for Non-IID Histopathological Image Classification

Cheng-Chang Tsai^{1,2}, Kai-Wen Cheng¹, Chun-Shien Lu¹

¹Institute of Information Science, Academia Sinica, Taipei, Taiwan, ROC

²Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, ROC
cctsa831@gmail.com, kevin0606@iis.sinica.edu.tw, lcs@iis.sinica.edu.tw

Abstract

Federated learning (FL) has shown success in collaboratively training a model among decentralized data resources without directly sharing privacy-sensitive training data. Despite recent advances, non-IID (non-independent and identically distributed) data poses an inevitable challenge that hinders the use of FL. In this work, we address the issue of non-IID histopathological images with feature distribution shifts from an intuitive perspective that has only received limited attention. Specifically, we address this issue from the perspective of data distribution by solely adjusting the data distributions of all clients. Building on the success of diffusion models in fitting data distributions and leveraging stain separation to extract the pivotal features that are closely related to the non-IID properties of histopathological images, we propose a **Federated Stain Distribution Alignment (FedSDA)** method. FedSDA aligns the stain distribution of each client with a target distribution in an FL framework to mitigate distribution shifts among clients. Furthermore, considering that training diffusion models on raw data in FL has been shown to be susceptible to privacy leakage risks, we circumvent this problem while still effectively achieving alignment. Extensive experimental results show that FedSDA is not only effective in improving baselines that focus on mitigating disparities across clients' model updates but also outperforms baselines that address the non-IID data issues from the perspective of data distribution. We show that FedSDA provides valuable and practical insights for the computational pathology community.

Introduction

Background

In recent years, advances in machine learning (ML) have significantly improved the performance of computer-aided diagnosis (CAD) systems (Chan, Hadjiiski, and Samala 2020; Maleki, Raahemi, and Nasiri 2023; Majumdar, Pramanik, and Sarkar 2023; Tsai, Chen, and Lu 2025; Raswa, Lu, and Wang 2025), which aid clinicians in analyzing medical images (e.g., histopathological images) and making more accurate and efficient decisions. Histopathological images play an indispensable role in the CAD of cancerous diseases (Mosquera-Lopez et al. 2015; Sun et al. 2020; Chen et al. 2017). However, due to noise, missing values, rarity of certain diseases, cost issues, and privacy concerns, the available data in a single medical center for training an ML model is often limited, thereby affecting the performance of

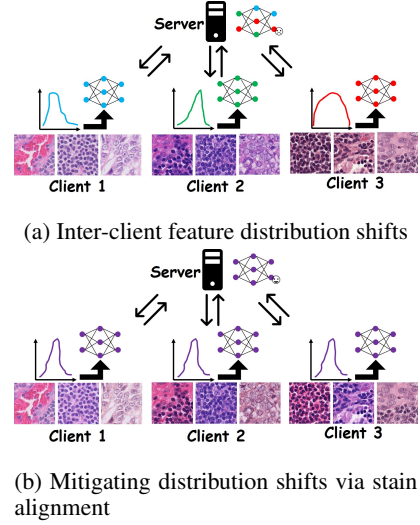


Figure 1: Illustration of two scenarios in FL. A model is collaboratively trained on histopathological images from multiple medical centers in a federated learning setting.

the trained model. To enhance the potential of data-driven ML models, collecting data from multiple medical centers is an intuitive and practical approach to increase the volume of data available for training models. However, the sensitive nature of medical images hinders the adoption of this approach.

To address this challenge and facilitate training, federated learning (FL) (McMahan et al. 2017), an emerging collaborative training framework that enables training on decentralized data from multiple medical centers without directly sharing sensitive training data, demonstrates great promise in addressing this hindrance. Despite the success of FL in histopathological image analysis (Lu et al. 2022; Adnan et al. 2022), non-IID (non-independent and identically distributed) data among clients still significantly deteriorates the performance of FL systems (Kairouz et al. 2021). This issue is particularly prevalent in FL settings that involve training models on histopathological image data from multiple medical centers (Bándi et al. 2019), as depicted in Fig. 1a. Therefore, this issue warrants urgent attention. In

this work, we focus on non-IID data with feature distribution shifts (unless otherwise specified, non-IID data and distribution shifts refer to non-IID histopathological image data and feature distribution shifts, respectively) (Li et al. 2021; Zhou and Konukoglu 2023; Qu et al. 2022) rather than on non-IID data with label distribution shifts (Deng, Luo, and Chen 2023; Wu et al. 2023).

Here is a commonly encountered example in a real medical center setting. Consider a model trained on data from one medical center to detect cancerous lesions in histopathological images. However, when deployed in another medical center, the model struggles due to differences in staining protocols and image acquisition scanners, which cause substantial chromatic variations across histopathological images (Gupta et al. 2020; Tellez et al. 2019). Even within the same medical center, it is common for a scanner of a different brand to be purchased, which can cause the model to not perform well in the histopathological images produced by the new scanner. This poses a significant challenge that hinders AI models from being portable and widely adopted in practical use.

Motivation

Existing works primarily address non-IID data issues in FL from the following two perspectives: (i) Optimization (Li et al. 2020; Acar et al. 2021; Gao et al. 2022; Qu et al. 2022; Fan et al. 2024; Dai et al. 2023; Karimireddy et al. 2020) and (ii) Normalization (Li et al. 2021; Zhang et al. 2023; Kang et al. 2024; Wagner et al. 2022; Shen et al. 2022). Despite these efforts, little attention has been paid to addressing non-IID data issues in FL solely from the perspective of data distribution (i.e., adjusting the data distributions of all clients to mitigate non-IID data issues, as depicted in Fig. 1b). Among the limited works in this direction, CCST (Chen et al. 2023) is the one specifically proposed to address non-IID data issues in FL by aligning each client’s data distribution with those of all the other.

However, we have observed that CCST (Chen et al. 2023) cannot preserve the structural information of histopathological images well, as shown in Figs. 17 and 18 (in the Appendix). This motivates us to design a better method that minimizes the loss of structural information. Since the core issue of non-IID data lies in stain variations, a promising approach is to separate histopathological images into the part containing stain-related information and the part containing structural information and to modify only the stain-related part while keeping the structural part unchanged. To achieve such separation, we adopt stain separation (Vahadane et al. 2016). Then, we adjust the distributions of the stain matrices to align the data distributions with each other, thereby addressing the non-IID data issues. However, in FL, stain separation (Vahadane et al. 2016) cannot reconstruct histopathological images for alignment without access to the stains from the other clients. To address this issue, we adopt a diffusion model (Ho, Jain, and Abbeel 2020) trained using FedAvg (McMahan et al. 2017) to fit the data distribution collectively formed by all clients (i.e., target distribution). Although this approach can achieve the goal, recent studies (Zhu et al. 2025; Gan, Miao, and Yang 2024) have shown

that training diffusion models on raw data via FedAvg carries a high risk of privacy leakage. For example, communicated gradients have been shown to leak private data under certain training configurations (Zhu, Liu, and Han 2019). Moreover, FedMIA (Zhu et al. 2025) can effectively determine whether a specific data sample belongs to the private dataset. However, among these threats, DataSteal (Gan, Miao, and Yang 2024) is a method specifically designed for diffusion models trained via federated learning and demonstrates superior attack performance. To mitigate these privacy risks, we train the diffusion model on stain matrices rather than raw data.

With all these concerns adequately addressed, we propose **Federated Stain Distribution Alignment (FedSDA)**, a method that aligns the stain distributions of all clients in an FL system with each other, as depicted in Fig. 1b. It is noteworthy that FedSDA shares a similar goal with normalization-based methods (Jiang, Wang, and Dou 2022; Li et al. 2021). However, FedSDA has the advantage of being aware of the structure of histopathological images and the stain distribution of data, but normalization-based methods do not pay special attention to these aspects. Furthermore, FedSDA does not require additional processing during testing, whereas normalization-based methods do.

Contributions

Our contributions are summarized as follows:

- We propose a federated stain distribution alignment method that mitigates feature distribution shifts using a diffusion model trained via federated learning, enabling each client to access the stains of all the other clients.
- The proposed method circumvents the need to train the diffusion model on raw data while still effectively achieving alignment, as direct training on raw data in FL has been shown to be susceptible to privacy leakage risks.
- Extensive experimental comparisons demonstrate that the proposed method not only improves the classification performance of the baselines (Li et al. 2020; Qu et al. 2022; Fan et al. 2024) that focus on mitigating disparities across clients’ model updates but also outperforms those (Chen et al. 2023; Jiang, Wang, and Dou 2022; Zhou and Konukoglu 2023) that address non-IID data issues from the perspective of data distribution.

Related Work

Federated Learning

FedAvg (McMahan et al. 2017) is a seminal work in FL and can be regarded as the foundation of all subsequent works.

Optimization Perspective The methods reviewed here share a similar design rationale to address non-IID data issues by mitigating disparities among updates. FedProx (Li et al. 2020) incorporates a proximal term into the loss function during local training to reduce the discrepancy between local and global models. SCAFFORD (Karimireddy et al. 2020) corrects local updates by estimating the difference between global and local update directions. FedDyn (Acar et al. 2021) dynamically updates each client’s regularizer

in each round to align the local solution with the solution of the global loss function in the limit. FedDC (Gao et al. 2022) introduces an auxiliary variable in local training to track the gap between local and global models. FedSAM (Qu et al. 2022) leverages the Sharpness Aware Minimization (SAM) (Foret et al. 2020) local optimizer to enhance local learning generalization. To further improve FedSAM (Qu et al. 2022), FedLESAM (Fan et al. 2024) locally estimates the global perturbation direction on the client side by computing the difference between global models received in the previous active round and the current round. FedGAMMA (Dai et al. 2023) incorporates FedSAM (Qu et al. 2022) with the correction technique proposed by SCAFFORD (Karimireddy et al. 2020) to mitigate disparities across all client updates.

Normalization Perspective In contrast, the methods reviewed here employ various normalization techniques to mitigate distribution shifts. However, these normalization techniques often induce the data to conform to a single feature rather than retaining the original distribution of the data (e.g., in Figs. 17 and 18 in the Appendix, the images in the bottom row are normalized to a similar stain, whereas those in the second row from the top exhibit more diverse stains). This highlights the fundamental difference between the proposed method and normalization-based methods. FedBN (Li et al. 2021), built on FedAvg (McMahan et al. 2017), uses batch normalization (BN) layers (Ioffe 2015) locally and excludes the parameters of the BN layers from aggregation on the central server. Zhang et al. study the role of layer normalization (LN) (Ba 2016) on non-IID data with label distribution shifts in FL, and their results verify that applying normalization to the feature from a classifier is the essence of LN, which improves classification performance under extreme label distribution shifts. FedNN (Kang et al. 2024) uses weight normalization (WN) (Salimans and Kingma 2016) and adaptive group normalization (AGN) to mitigate variations in model weights and features, respectively, in concept drift data (i.e., feature distribution shifts). Finally, HarmoFL (Jiang, Wang, and Dou 2022) employs SAM (Foret et al. 2020) and amplitude normalization to mitigate local and global drift, respectively.

Distribution Shifts

Non-IID data poses a significant issue in FL, also known as distribution shifts (Gao et al. 2023; Tsai et al. 2024; Zhang et al. 2021; Koh et al. 2021; Raswa, Lu, and Wang 2025), where gaps exist among data distributions. Many previous works (Gao et al. 2023; Tsai et al. 2024; Nie et al. 2022; Tsai, Chen, and Lu 2025; Hoffman et al. 2018) have been proposed to address this issue from the perspective of domain adaptation using generative models. Specifically, they adapt data from one dataset (i.e., target domain) to match the distribution of another dataset (i.e., source domain). CyCADA (Hoffman et al. 2018), based on CycleGAN (Zhu et al. 2017), adapts data by translating between domains, guided by a specific discriminatively trained task. DiffPure (Nie et al. 2022) uses diffusion models (Ho, Jain, and Abbeel 2020) to purify adversarial images (Goodfel-

low, Shlens, and Szegedy 2014) (i.e., images from target domain), generating purified images that closely follow the distribution of clean images (i.e., images from source domain). DDA (Gao et al. 2023) also uses diffusion models (Ho, Jain, and Abbeel 2020) to adapt images from a target domain (or possibly multiple target domains) to a source domain without significantly losing class information, considering the distance between the reference and noisy images after applying a low-pass filter. GDA (Tsai et al. 2024) refines DDA (Gao et al. 2023) by adopting the structural guidance, which is defined by a marginal entropy loss derived from the classifier, along with style and content preservation losses. TT-SaD (Tsai, Chen, and Lu 2025) adapts the stains of histopathological images using diffusion models (Ho, Jain, and Abbeel 2020). In addition to these domain adaptation methods that are not designed for FL, Chen et al. propose CCST (Chen et al. 2023), which solely adjusts the data distributions of all clients by augmenting the data in each client through cross-client style transfer based on adaptive instance normalization (AdaIN) (Huang and Belongie 2017) for federated domain generalization.

Preliminaries

Notations

In an FL system of K clients, the i -th client, denoted by C_i , only has access to histopathological images $\mathcal{X}_i = \{\mathbf{x}_{i,j}\}_{j=1}^{n_i}$ and the corresponding labels $\mathcal{Y}_i = \{\mathbf{y}_{i,j}\}_{j=1}^{n_i}$, and they constitute a dataset $\mathcal{D}_i = \{(\mathbf{x}_{i,j}, \mathbf{y}_{i,j})\}_{j=1}^{n_i}$, where n_i denotes the number of data in C_i , for $i \in \{1, 2, \dots, K\}$. In addition, the data distribution of C_i is denoted by $\mathcal{P}_i$, and all data pairs $(\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{D}_i$ are drawn from $\mathcal{P}_i$, i.e., $(\mathbf{x}_i, \mathbf{y}_i) \sim \mathcal{P}_i$. Finally, let $\mathbb{E}$, $\log(\cdot)$, $\|\cdot\|_F$, $\|\cdot\|_1$, $\|\cdot\|_2$, $\mathbf{x}(:, i)$, and $\exp(\cdot)$ denote the expectation, the element-wise logarithmic function, the Frobenius norm, the ℓ_1 norm, the ℓ_2 norm, the i -th column of $\mathbf{x}$, and the element-wise exponential function, respectively. We will frequently omit subscripts for simplicity.

Problem Formulation

The goal of FL is to train a model f_θ , parameterized by θ , on decentralized data without directly sharing privacy-sensitive training data. Specifically, the objective can be formulated as follows (McMahan et al. 2017; Jiang et al. 2021):

$$\arg \min_{\theta} \frac{1}{K} \sum_{i=1}^K \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \in \mathcal{D}_i} [\ell(f_\theta(\mathbf{x}), \mathbf{y})], \quad (1)$$

where ℓ denotes a task-dependent loss function, e.g., cross-entropy loss for a classification task.

In this work, we consider solving Eq. (1) under inter-client feature distribution shifts (Tan et al. 2024), which are defined as follows:

$$\mathcal{P}_i(\mathbf{x}|\mathbf{y}) \neq \mathcal{P}_j(\mathbf{x}|\mathbf{y}) \quad \text{for all } i \neq j. \quad (2)$$

This scenario is prevalent in histopathological image analysis (Bándi et al. 2019; Wilm et al. 2023), as depicted in Fig. 1a.

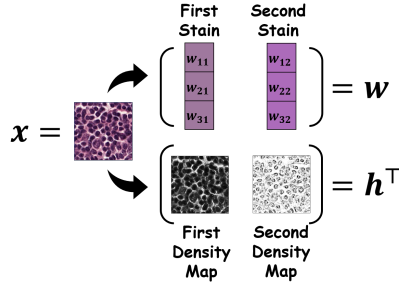


Figure 2: Illustration of stain separation. Given an H&E-stained histopathological image $\mathbf{x}$, we can decompose it into a stain matrix $\mathbf{w}$ of size 3×2 and a stain density map $\mathbf{h}$ of size $2 \times N$, where r represents the number of stains (in this case, $r = 2$). Each w_{ij} represents an element of $\mathbf{w}$.

Stain Separation

Stain separation (Vahadane et al. 2016), as depicted in Fig. 2, aims to decompose a histopathological image $\mathbf{x} \in \mathbb{R}^{m \times N}$, where $m = 3$ denotes the RGB channels and N is the number of pixels in $\mathbf{x}$, into a stain matrix, denoted by $\mathbf{w} \in \mathbb{R}^{m \times r}$, and the corresponding stain density map, denoted by $\mathbf{h} \in \mathbb{R}^{r \times N}$, so that $\mathbf{x}$ can be reconstructed by $\mathbf{w}$ and $\mathbf{h}$. Note that the columns of $\mathbf{w}$ and the rows of $\mathbf{h}$ represent the chromatic basis of each stain and the concentration of each stain, respectively. For more details on stain separation, please refer to (Vahadane et al. 2016).

Stain separation is formulated as the following optimization problem:

$$\begin{aligned} \arg \min_{\mathbf{w}, \mathbf{h}} \quad & \frac{1}{2} \left\| -\log \frac{\mathbf{x}}{I_0} - \mathbf{w}\mathbf{h} \right\|_F^2 + \lambda \|\mathbf{h}\|_1 \\ \text{s.t.} \quad & \mathbf{w}, \mathbf{h} \geq 0 \quad \text{and} \quad \|\mathbf{w}(:, i)\|_2^2 = 1 \quad \forall i, \end{aligned} \quad (3)$$

where λ is a regularization parameter that balances the regularization and loss terms, and I_0 is the illuminating light intensity of an image (typically 255 for 8-bit images). This optimization problem is solved iteratively by optimizing one set of parameters (e.g., $\mathbf{w}$) while keeping the other (i.e., $\mathbf{h}$) fixed. After obtaining the stain matrix $\mathbf{w}$ and the stain density map $\mathbf{h}$ of a histopathological image $\mathbf{x}$, the reconstructed image $\hat{\mathbf{x}}$ is computed as:

$$\hat{\mathbf{x}} = I_0 \exp(-\mathbf{w}\mathbf{h}). \quad (4)$$

For clarity, we denote Eq. (3) and Eq. (4) as $\mathcal{S}$ and $\mathcal{R}$, respectively, which leads to the following two expressions:

$$\{\mathbf{w}, \mathbf{h}\} = \mathcal{S}(\mathbf{x}) \quad \text{and} \quad \hat{\mathbf{x}} = \mathcal{R}(\mathbf{w}, \mathbf{h}). \quad (5)$$

Diffusion Models

Diffusion models (Ho, Jain, and Abbeel 2020) are known to fit a given data distribution, denoted by $q(\mathbf{x}_0)$, so that the generated data belong to $q(\mathbf{x}_0)$, and consist of two processes (i.e., the forward and reverse processes). Given $\mathbf{x}_0 \sim q(\mathbf{x}_0)$, the forward process gradually adds Gaussian noise to $\mathbf{x}_0$ to produce latent samples $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$, where T is a predefined constant, via

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) := \mathcal{N}\left(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}\right), \quad (6)$$

where $\{\beta_t\}_{t=1}^T$ is a variance schedule, $\mathcal{N}$ denotes a Gaussian distribution, and $\mathbf{I}$ denotes the identity matrix. Conversely, the reverse process is formulated as:

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) := \mathcal{N}\left(\mathbf{x}_{t-1}; \frac{\mathbf{x}_t}{\sqrt{\alpha_t}} - \frac{\beta_t \epsilon_\theta(\mathbf{x}_t, t)}{\sqrt{\alpha_t(1 - \bar{\alpha}_t)}}, \beta_t \mathbf{I}\right), \quad (7)$$

where $\alpha_t := 1 - \beta_t$, $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$, and ϵ_θ is a neural network parameterized by θ . Finally, to train a conditional diffusion model, we optimize the following problem:

$$\min_{\theta} \mathbb{E}_{t, \mathbf{x}_0, c, \epsilon} \|\epsilon - \epsilon_\theta(\mathbf{x}_t, t, c)\|_2^2, \quad (8)$$

where t is the timestep, c is the condition, and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$.

Method

Overview of FedSDA

To mitigate the issues defined in Eq. (2), FedSDA, which is proposed to align the stain distributions of all clients with a target distribution, consists of two steps: (i) Target distribution fitting and (ii) Stain alignment. The first step is to use a diffusion model to fit the stain distribution formed collectively by all clients (i.e., target distribution). This diffusion model provides each client with access to the stain matrices of the other clients, which is a necessary component for the second step. With the diffusion model, the second step can be applied to histopathological images of each client, thereby achieving the goal of addressing the non-IID data issues. We provide the workflow of FedSDA in Fig. 3.

Target Distribution Fitting

FedSDA requires a generative model to generate the stain matrices of the other clients for the second step of FedSDA (i.e., stain alignment). Due to the recent success of diffusion models in generating synthetic data and modeling complex distributions (Kotelnikov et al. 2023; Dhariwal and Nichol 2021; Ho et al. 2022), we adopt a diffusion model as the generative model for FedSDA.

Here, we introduce the diffusion model adopted in FedSDA. Since the model is trained on low-dimensional data, we adopt a single-layer transformer model as the backbone of the diffusion model, denoted by $\mathcal{G}$. This lightweight network architecture ensures that the additional overhead is minimal. This model takes the timestep t , the condition c , and the noisy input $\mathbf{w}_t$ as inputs, and estimates the noise added to $\mathbf{w}_t$. Since the noisy data $\mathbf{w}_t$ fed into $\mathcal{G}$ is a stain matrix with added noise rather than a sequence of tokens, it is necessary to adjust the way $\mathbf{w}_t$ is processed by $\mathcal{G}$. Specifically, we need to determine how $\mathbf{w}_t$ is represented as tokens, as $\mathcal{G}$ treats all inputs, including t , c , and $\mathbf{w}_t$, as tokens. Intuitively, we treat each element of $\mathbf{w}_t$ as a token and concatenate (t, c) with the tokens of $\mathbf{w}_t$ to form the input of $\mathcal{G}$, as depicted in Fig. 4.

To train a diffusion model on stain matrices from all clients without directly sharing them, we follow the previous works (Huang et al. 2024; Tun et al. 2023; Vora et al. 2024; de Goede, Cox, and Decouchant 2024) and adopt FedAvg (McMahan et al. 2017). Specifically, in each communication round, each client downloads the diffusion model

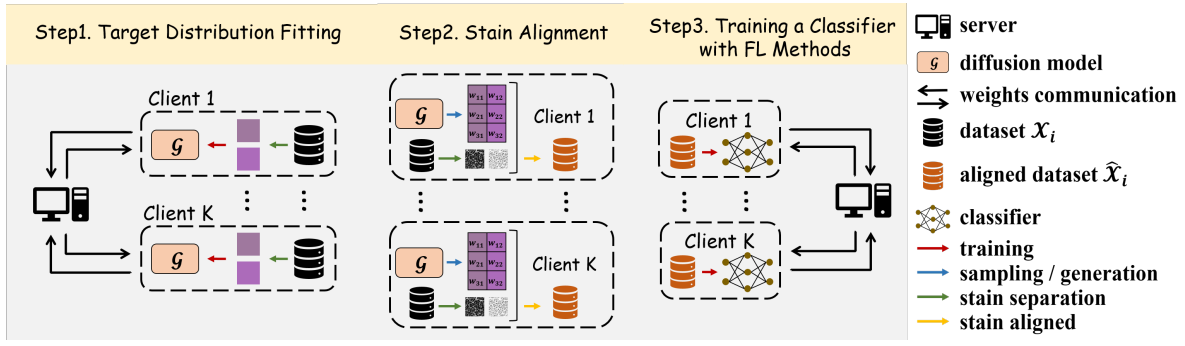


Figure 3: Workflow of our FedSDA method

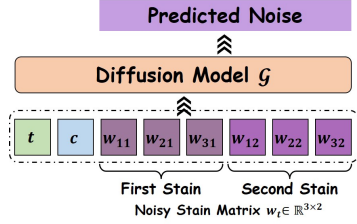


Figure 4: Illustration of how w_t is fed into G .

weights from the central server, then trains the model locally, and finally uploads the local weights to the central server for aggregation. We provide the pseudocode in Algorithm 1 (in the Appendix) for this training process. Note that the diffusion model we train is conditional, meaning that each stain matrix requires a condition c to form an input for G . For client C_i , we use its index i as the condition c for the corresponding stain matrices. For ease of reference, we denote the stain matrix generated under a given condition c as $G(c)$. The condition used here determines the client to which the stain matrices belong and corresponds to the client index. Therefore, to generate the stain matrices belonging to client C_i , we apply $G(i)$.

Stain Alignment

In the first step of FedSDA, the central server ends up with a conditional diffusion model G , trained on the stain matrices from all clients in an FL system. This diffusion model G is used in the second step of FedSDA to sample stain matrices belonging to other clients for stain alignment. With access to other clients' stain matrices, we can align the data distribution of client C_i with some distribution $\hat{P}$ (ideally, it is the data distribution considering all clients' data as a whole), i.e., $P_i(\mathbf{x}|\mathbf{y}) \approx \hat{P}(\mathbf{x}|\mathbf{y})$. As a result of aligning through all clients, we can obtain the following results for clients C_i and C_j :

$$P_i(\mathbf{x}|\mathbf{y}) \approx P_j(\mathbf{x}|\mathbf{y}) \quad \text{for all } i \neq j, \quad (9)$$

which mitigates the non-IID characteristic of data across different clients. Note that the reason we can achieve Eq. (9) is that we have identified stain as the essential factor that causes feature distribution shifts. Therefore, reducing stain distribution shifts can achieve Eq. (9).

Specifically, the diffusion model G is downloaded by all clients from the central server. Since all clients follow the same procedure (denoted by $\mathcal{A}$ hereafter) to achieve Eq. (9), we present the procedure $\mathcal{A}$ for a given client C_i as an example for simplicity, and the same procedure applies to all other clients. We provide the pseudocode in Algorithm 2 (in the Appendix).

Given a dataset $\mathcal{D}_i$ of client C_i , due to the label-free nature of FedSDA, $\mathcal{A}$ only requires the input dataset $\mathcal{X}_i$ to obtain an aligned dataset $\hat{\mathcal{X}}_i$, i.e., $\hat{\mathcal{X}}_i = \mathcal{A}(\mathcal{X}_i)$. To balance the stains from each client in $\hat{\mathcal{X}}_i$, we randomly partition $\mathcal{X}_i$ into K (i.e., the number of clients) subsets as $\mathcal{X}_i = \{\mathcal{X}_{i,1}, \mathcal{X}_{i,2}, \dots, \mathcal{X}_{i,K}\}$, and each subset contains the same number of samples (denoted by $n = n_i/K$). For each subset indexed by j ($1 \leq j \leq K$), we use $\mathcal{S}$ (in Eq. (5)) to obtain the corresponding stain density maps (denoted by $\{\mathbf{h}_{i,j,k}\}_{k=1}^n$ hereafter) from the input data (denoted by $\{\mathbf{x}_{i,j,k}\}_{k=1}^n$ hereafter) by applying $\mathcal{S}(\mathbf{x}_{i,j,k})$ for each k , where $\mathcal{X}_{i,j} = \{\mathbf{x}_{i,j,k}\}_{k=1}^n$. Then, $\mathcal{R}$ (in Eq. (5)) is applied to reconstruct a dataset with the stain of client C_j . Specifically, for each $k \in \{1, 2, \dots, n\}$, the reconstructed $\hat{\mathbf{x}}_{i,j,k}$ is obtained by $\mathcal{R}(G(j), \mathbf{h}_{i,j,k})$, i.e., $\hat{\mathbf{x}}_{i,j,k} = \mathcal{R}(G(j), \mathbf{h}_{i,j,k})$. Thus, we obtain the reconstructed dataset $\hat{\mathcal{X}}_{i,j} = \{\hat{\mathbf{x}}_{i,j,k}\}_{k=1}^n$, and further obtain the aligned dataset $\hat{\mathcal{X}}_i = \{\hat{\mathcal{X}}_{i,1}, \hat{\mathcal{X}}_{i,2}, \dots, \hat{\mathcal{X}}_{i,K}\}$.

So far, we have introduced the procedure of stain alignment (i.e., $\mathcal{A}$). After all clients have applied $\mathcal{A}$ to their datasets, a tumor classifier can be trained in the FL system (i.e., the third step in Fig. 3).

Experiments

Experimental Settings

Building Block	Hidden size	#Heads	#Params	FD
MLP	32	—	4.7K	18.91
Transformer	32	8	13.2K	0.40

Table 1: Configuration details of diffusion models with different types of building blocks, and comparison of their FD scores. Note that both models have a single layer.

Datasets We evaluated our proposed method on three datasets, including Mitos & Atypia 14 (MA14) (Roux 2014), CAMELYON17 (C17) (Bándi et al. 2019), and AGGC22 (A22) (Huo et al. 2024). Each dataset contains whole-slide images (WSIs) stained with hematoxylin and eosin (H&E) dyes, so the number of stains is 2. Since we focus on patch-level classification in this work, we followed (Guo et al. 2019) to extract histopathological image patches from each WSI. For more detailed information on each dataset, please refer to Table 4, 5, and 6 (in the Appendix). In addition, we provide a few sample images of each dataset in Figs. 8, 9, and 10 (in the Appendix).

Baselines We adopted FedAvg (McMahan et al. 2017), FedProx (Li et al. 2020), FedSAM (Qu et al. 2022), FedLESAM (Fan et al. 2024), and FedFA (Zhou and Konukoglu 2023) as baselines to evaluate performance before and after applying FedSDA. Additionally, we compared the performance with CCST (Chen et al. 2023) and the amplitude normalization in HarmoFL (Jiang, Wang, and Dou 2022) (denoted by amp-norm), since these methods address non-IID data issues solely from the data distribution perspective, similar to FedSDA. For a clearer understanding of amp-norm, please refer to the pseudocode in Algorithm 3 (in the Appendix).

Evaluation Metrics Due to the class-imbalanced nature of the A22 dataset, we evaluated tumor classification performance using the area under the receiver operating characteristic curve (AUROC) (Hanley and McNeil 1982) and the area under the precision-recall curve (AUPRC) (Boyd, Eng, and Page 2013). For the class-balanced C17 dataset, we used AUROC only. On the other hand, following (Tsai, Chen, and Lu 2025; Zingman et al. 2024), we adopted the Structural Similarity Index (SSIM) (Wang et al. 2004), Wasserstein Distance (WD) (Ramdas, García Trillos, and Cuturi 2017), Fréchet Inception Distance (FID) (Heusel et al. 2017), and Kernel Inception Distance (KID) (Bińkowski et al. 2018) to assess image quality. SSIM measured the loss of structural information, which is important for tumor classification. WD computed the discrepancy between two images in terms of chromatic appearance and was adopted because stains are a pivotal factor in color variations in histopathological images. Both FID and KID (an improved version of FID) were adopted to measure the distance between two data distributions.

Implementation Details Each domain in a dataset is considered a client, where each domain refers to a subset with similar stains from the same hospital. We trained the diffusion model used in FedSDA with 3 communication rounds, 300 local epochs, a batch size of 65,536, and AdamW optimizer (Loshchilov 2017) with a weight decay of $3e^{-2}$ and a learning rate of $2e^{-4}$. We provide the configuration of the single-layer transformer model that we used in Table 1. For tumor classification, we adopted DenseNet-121 (Huang et al. 2017) as the classifier due to its wide adoption for histopathological images. We trained classifiers on the C17 and A22 datasets with 100 communication rounds, one local epoch, a batch size of 128, cross-entropy loss, and SGD

optimizer with a learning rate of $1e^{-3}$. We implemented all the methods using PyTorch and conducted the experiments on an NVIDIA Tesla V100 GPU.

Method	C17	A22	
	AUROC↑	AUROC↑	AUPRC↑
FedAvg w/ FedSDA	90.97 93.84	78.99 90.38	46.60 64.20
FedProx w/ FedSDA	90.80 93.89	79.82 90.44	45.71 64.89
FedSAM w/ FedSDA	90.87 93.04	82.17 92.09	48.21 67.45
FedLESAM w/ FedSDA	90.17 92.68	79.42 90.20	44.39 64.34
FedFA w/ FedSDA	92.31 94.92	82.62 91.79	48.45 66.86

Table 2: Results of tumor classification before and after applying FedSDA.

Main Results

Evaluation of Tumor Classification First, we applied FedSDA to the five baselines (i.e., FedAvg (McMahan et al. 2017), FedProx (Li et al. 2020), FedSAM (Qu et al. 2022), FedLESAM (Fan et al. 2024), and FedFA (Zhou and Konukoglu 2023)) to verify whether the performance can be improved with the use of FedSDA. Table 2 demonstrates that incorporating FedSDA yields a significant performance improvement across the baselines. Note that FedAvg combined with FedSDA can outperform the other four baselines by a significant margin. Additionally, Fig. 5 depicts the convergence curves corresponding to the results in Table 2 and demonstrates that incorporating FedSDA enables convergence not only to a more optimal solution but also at a faster rate. Second, the results of different modules embedded in FedAvg (McMahan et al. 2017) are compared in Table 3, and it can be observed that FedSDA significantly

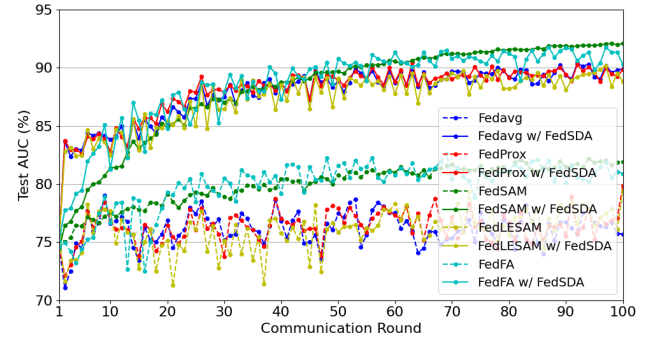


Figure 5: Test AUC vs. communication rounds on the A22 dataset before and after applying FedSDA.

Method	C17	A22		MA14 & C17 & A22			
	AUROC↑	AUROC↑	AUPRC↑	SSIM↑	FID↓	KID↓	WD↓
FedAvg	90.97	78.99	46.60	1.0000	55.092	0.1986	0.0013
w/ CCST	71.89	80.09	38.43	0.2462	20.777	0.0599	0.0007
w/ amp-norm	77.49	77.01	42.83	0.9731	54.774	0.1956	0.0012
w/ FedSDA	93.84	90.38	64.20	0.9969	40.311	0.1365	0.0009

Table 3: Comparison of tumor classification and image quality (averaged over three datasets, i.e., MA14, C17, and A22) across different modules embedded in FedAvg.

outperforms the other modules, regardless of whether the dataset is imbalanced.

Evaluation of Image Quality Table 3 also shows a comparison of image quality across different embedded modules. It can be observed that FedSDA outperforms both the baseline without FedSDA and the one using amp-norm. We also observe that CCST (Chen et al. 2023) outperforms FedSDA in terms of FID, KID, and WD; however, it loses significantly more structural information, as indicated by a drop in SSIM to 0.2462, which substantially compromises its effectiveness for tumor classification. This is clearly evidenced in Table 3. Finally, we visualize histopathological images before and after applying FedSDA in Fig. 6. After applying FedSDA, the low-dimensional features of each client’s data tend to be distributed more similarly than before. Additional visualization results are provided in the Appendix. Specifically, we include visualizations of the stain matrices (Figs. 14, 15, and 16 in the Appendix), comprising both those decomposed from the histopathological images of each client and those generated by the diffusion model.

Ablation Study on Diffusion Model in FedSDA

The ablation studies conducted here are measured and compared in terms of the Fréchet Distance (FD) between two sets of stain matrices.

Communication Rounds vs. Local Epochs To examine the effect of the number of local epochs on communication rounds, we set the number of local epochs as 1, 2, 3, 4, 100, 200, and 300, and recorded the corresponding number of communication rounds required for convergence. As

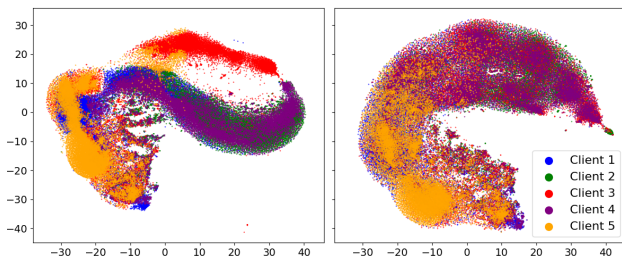


Figure 6: Visualizations of the C17 dataset before and after applying FedSDA by t-SNE. Please refer to Figs. 11, 12, and 13 (in the Appendix) for detailed results on all datasets.

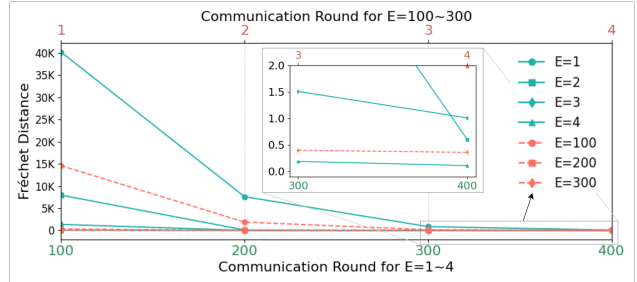


Figure 7: Fréchet distance vs. communication rounds under different numbers of local epochs on the C17 dataset.

shown in Fig. 7, many more communication rounds are required for convergence when using fewer local epochs (i.e., 1 – 4). However, training converges with an FD of 0.40 and a WD of 0.07 in only 3 communication rounds when the number of local epochs was set to 300. This observation suggests that increasing the number of local epochs can significantly reduce communication costs while maintaining model performance.

Network Architecture With the number of communication rounds and local epochs set to 3 and 300, respectively, we compared the performance of a single-layer transformer model with a single-layer MLP model. As shown in Table 1, a much more lightweight network architecture with MLP as the main building block significantly deteriorates performance.

Conclusion

In this paper, we propose FedSDA to address the challenge of non-IID histopathological images with feature distribution shifts in a federated environment. By leveraging stain separation and fitting a target distribution using a diffusion model without directly training on raw data, FedSDA mitigates stain distribution shifts across clients while preventing raw data from being directly exposed to the risks of leakage in a federated environment. Extensive experiments show that FedSDA not only improves the performance of optimization-based methods but also outperforms distribution-based methods. This work highlights the potential of focusing on data-centric solutions within FL, particularly for applications involving complex, non-IID medical image datasets.

Acknowledgments

This work was supported by the National Science and Technology Council (NSTC), Taiwan, ROC, under Grants NSTC 113-2634-F-006-002 and 114-2221-E-001 -010 -MY2.

References

- Acar, D. A. E.; Zhao, Y.; Navarro, R. M.; Mattina, M.; Whatmough, P. N.; and Saligrama, V. 2021. Federated learning based on dynamic regularization. *arXiv preprint arXiv:2111.04263*.
- Adnan, M.; Kalra, S.; Cresswell, J. C.; Taylor, G. W.; and Tizhoosh, H. R. 2022. Federated learning and differential privacy for medical image analysis. *Scientific reports*, 12(1): 1953.
- Ba, J. L. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450*.
- Bińkowski, M.; Sutherland, D. J.; Arbel, M.; and Gretton, A. 2018. Demystifying MMD GANs. In *International Conference on Learning Representations*.
- Boyd, K.; Eng, K. H.; and Page, C. D. 2013. Area under the precision-recall curve: point estimates and confidence intervals. In *In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD*, 451–466.
- Bándi, P.; Geessink, O.; Manson, Q.; Van Dijk, M.; Balkenhol, M.; Hermesen, M.; Ehteshami Bejnordi, B.; Lee, B.; Paeng, K.; Zhong, A.; Li, Q.; Zanjani, F. G.; Zinger, S.; Fukuta, K.; Komura, D.; Ovtcharov, V.; Cheng, S.; Zeng, S.; Thagaard, J.; Dahl, A. B.; Lin, H.; Chen, H.; Jacobsson, L.; Hedlund, M.; Çetin, M.; Halıcı, E.; Jackson, H.; Chen, R.; Both, F.; Franke, J.; Küsters-Vandeveld, H.; Vreuls, W.; Bult, P.; van Ginneken, B.; van der Laak, J.; and Litjens, G. 2019. 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.
- Chan, H.-P.; Hadjiiski, L. M.; and Samala, R. K. 2020. Computer-aided diagnosis in the era of deep learning. *Medical physics*, 47(5): e218–e227.
- Chen, J.; Jiang, M.; Dou, Q.; and Chen, Q. 2023. Federated domain generalization for image recognition via cross-client style transfer. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 361–370.
- Chen, J.-M.; Li, Y.; Xu, J.; Gong, L.; Wang, L.-W.; Liu, W.-L.; and Liu, J. 2017. Computer-aided prognosis on breast cancer with hematoxylin and eosin histopathology images: A review. *Tumor Biology*, 39(3): 1010428317694550.
- Dai, R.; Yang, X.; Sun, Y.; Shen, L.; Tian, X.; Wang, M.; and Zhang, Y. 2023. FedGAMMA: Federated Learning With Global Sharpness-Aware Minimization. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14.
- de Goede, M.; Cox, B.; and Decouchant, J. 2024. Training Diffusion Models with Federated Learning. *arXiv preprint arXiv:2406.12575*.
- Deng, Z.; Luo, L.; and Chen, H. 2023. Scale Federated Learning for Label Set Mismatch in Medical Image Classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 118–127. Springer.
- Dhariwal, P.; and Nichol, A. 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34: 8780–8794.
- Fan, Z.; Hu, S.; Yao, J.; Niu, G.; Zhang, Y.; Sugiyama, M.; and Wang, Y. 2024. Locally Estimated Global Perturbations are Better than Local Perturbations for Federated Sharpness-aware Minimization. In Salakhutdinov, R.; Kolter, Z.; Heller, K.; Weller, A.; Oliver, N.; Scarlett, J.; and Berkenkamp, F., eds., *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, 12858–12881. PMLR.
- Foret, P.; Kleiner, A.; Mobahi, H.; and Neyshabur, B. 2020. Sharpness-aware minimization for efficiently improving generalization. *arXiv preprint arXiv:2010.01412*.
- Gan, Y.; Miao, J.; and Yang, Y. 2024. DataStealing: Steal Data from Diffusion Models in Federated Learning with Multiple Trojans. *Advances in Neural Information Processing Systems*, 37: 132614–132646.
- Gao, J.; Zhang, J.; Liu, X.; Darrell, T.; Shelhamer, E.; and Wang, D. 2023. Back to the source: Diffusion-driven adaptation to test-time corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11786–11796.
- Gao, L.; Fu, H.; Li, L.; Chen, Y.; Xu, M.; and Xu, C.-Z. 2022. Feddc: Federated learning with non-iid data via local drift decoupling and correction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10112–10121.
- Goodfellow, I. J.; Shlens, J.; and Szegedy, C. 2014. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.
- Guo, Z.; Liu, H.; Ni, H.; Wang, X.; Su, M.; Guo, W.; Wang, K.; Jiang, T.; and Qian, Y. 2019. A fast and refined cancer regions segmentation framework in whole-slide breast pathological images. *Scientific reports*, 9(1): 882.
- Gupta, A.; Duggal, R.; Gehlot, S.; Gupta, R.; Mangal, A.; Kumar, L.; Thakkar, N.; and Satpathy, D. 2020. GCTI-SN: Geometry-inspired chemical and tissue invariant stain normalization of microscopic medical images. *Medical Image Analysis*, 65: 101788.
- Hanley, J. A.; and McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (roc) curve. In *Radiology*, 29–36.
- Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.
- Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.
- Ho, J.; Salimans, T.; Gritsenko, A.; Chan, W.; Norouzi, M.; and Fleet, D. J. 2022. Video diffusion models. *Advances in Neural Information Processing Systems*, 35: 8633–8646.

- Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.-Y.; Isola, P.; Saenko, K.; Efros, A.; and Darrell, T. 2018. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, 1989–1998. Pmlr.
- Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.
- Huang, J.; Hong, C.; Chen, L. Y.; and Roos, S. 2024. Gradient Inversion of Federated Diffusion Models. *arXiv preprint arXiv:2405.20380*.
- Huang, X.; and Belongie, S. 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, 1501–1510.
- Huo, X.; Ong, K. H.; Lau, K. W.; Gole, L.; Young, D. M.; Tan, C. L.; Zhu, X.; Zhang, C.; Zhang, Y.; Li, L.; et al. 2024. A comprehensive AI model development framework for consistent Gleason grading. *Communications Medicine*, 4(1): 84.
- Ioffe, S. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
- Jiang, M.; Li, X.; Zhang, X.; Kamp, M.; and Dou, Q. 2021. Unified: A unified framework for federated learning on non-iid image features. *arXiv preprint arXiv:2110.09974*.
- Jiang, M.; Wang, Z.; and Dou, Q. 2022. HarmoFL: Harmonizing Local and Global Drifts in Federated Learning on Heterogeneous Medical Images. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(1): 1087–1095.
- Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. 2021. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, 14(1–2): 1–210.
- Kang, M.; Kim, S.; Jin, K. H.; Adeli, E.; Pohl, K. M.; and Park, S. H. 2024. FedNN: Federated learning on concept drift data using weight and adaptive group normalizations. *Pattern Recognition*, 149: 110230.
- Karimireddy, S. P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; and Suresh, A. T. 2020. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, 5132–5143. PMLR.
- Koh, P. W.; Sagawa, S.; Marklund, H.; Xie, S. M.; Zhang, M.; Balsubramani, A.; Hu, W.; Yasunaga, M.; Phillips, R. L.; Gao, I.; et al. 2021. Wilds: A benchmark of in-the-wild distribution shifts. In *International conference on machine learning*, 5637–5664. PMLR.
- Kotelnikov, A.; Baranchuk, D.; Rubachev, I.; and Babenko, A. 2023. Tabddpm: Modelling tabular data with diffusion models. In *International Conference on Machine Learning*, 17564–17579. PMLR.
- Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2: 429–450.
- Li, X.; Jiang, M.; Zhang, X.; Kamp, M.; and Dou, Q. 2021. Fedbn: Federated learning on non-iid features via local batch normalization. *arXiv preprint arXiv:2102.07623*.
- Loshchilov, I. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- Lu, M. Y.; Chen, R. J.; Kong, D.; Lipkova, J.; Singh, R.; Williamson, D. F.; Chen, T. Y.; and Mahmood, F. 2022. Federated learning for computational pathology on gigapixel whole slide images. *Medical image analysis*, 76: 102298.
- Majumdar, S.; Pramanik, P.; and Sarkar, R. 2023. Gamma function based ensemble of CNN models for breast cancer detection in histopathology images. *Expert Systems with Applications*, 213: 119022.
- Maleki, A.; Raahemi, M.; and Nasiri, H. 2023. Breast cancer diagnosis from histopathology images using deep neural network and XGBoost. *Biomedical Signal Processing and Control*, 86: 105152.
- McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.
- Mosquera-Lopez, C.; Agaian, S.; Velez-Hoyos, A.; and Thompson, I. 2015. Computer-Aided Prostate Cancer Diagnosis From Digitized Histopathology: A Review on Texture-Based Systems. *IEEE Reviews in Biomedical Engineering*, 8: 98–113.
- Nie, W.; Guo, B.; Huang, Y.; Xiao, C.; Vahdat, A.; and Anandkumar, A. 2022. Diffusion Models for Adversarial Purification. In *International Conference on Machine Learning*, 16805–16827. PMLR.
- Qu, Z.; Li, X.; Duan, R.; Liu, Y.; Tang, B.; and Lu, Z. 2022. Generalized federated learning via sharpness aware minimization. In *International conference on machine learning*, 18250–18280. PMLR.
- Ramdas, A.; García Trillos, N.; and Cuturi, M. 2017. On wasserstein two-sample testing and related families of non-parametric tests. *Entropy*, 19(2): 47.
- Raswa, F. H.; Lu, C.-S.; and Wang, J.-C. 2025. HistoFS: Non-IID Histopathologic Whole Slide Image Classification via Federated Style Transfer with RoI-Preserving. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 30251–30260.
- Roux, L. 2014. Detection of mitosis and evaluation of nuclear atypia score in breast cancer histological images. In *22nd International Conference on Pattern Recognition, Stockholm, Sweden*.
- Salimans, T.; and Kingma, D. P. 2016. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. *Advances in neural information processing systems*, 29.
- Shen, Y.; Sowmya, A.; Luo, Y.; Liang, X.; Shen, D.; and Ke, J. 2022. A federated learning system for histopathology image analysis with an orchestral stain-normalization GAN. *IEEE Transactions on Medical Imaging*, 42(7): 1969–1981.
- Sun, H.; Zeng, X.; Xu, T.; Peng, G.; and Ma, Y. 2020. Computer-Aided Diagnosis in Histopathological Images of

- the Endometrium Using a Convolutional Neural Network and Attention Mechanisms. *IEEE Journal of Biomedical and Health Informatics*, 24(6): 1664–1676.
- Tan, Y.; Chen, C.; Zhuang, W.; Dong, X.; Lyu, L.; and Long, G. 2024. Is heterogeneity notorious? taming heterogeneity to handle test-time shift in federated learning. *Advances in Neural Information Processing Systems*, 36.
- Tellez, D.; Litjens, G.; Bándi, P.; Bulten, W.; Bokhorst, J.-M.; Ciompi, F.; and Van Der Laak, J. 2019. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical image analysis*, 58: 101544.
- Tsai, C.-C.; Chen, Y.-C.; and Lu, C.-S. 2025. Test-Time Stain Adaptation with Diffusion Models for Histopathology Image Classification. In *European Conference on Computer Vision*, 257–275. Springer.
- Tsai, Y.-Y.; Chen, F.-C.; Chen, A. Y.; Yang, J.; Su, C.-C.; Sun, M.; and Kuo, C.-H. 2024. GDA: Generalized Diffusion for Robust Test-time Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 23242–23251.
- Tun, Y. L.; Thwal, C. M.; Yoon, J. S.; Kang, S. M.; Zhang, C.; and Hong, C. S. 2023. Federated learning with diffusion models for privacy-sensitive vision tasks. In *2023 International Conference on Advanced Technologies for Communications (ATC)*, 305–310. IEEE.
- Vahadane, A.; Peng, T.; Sethi, A.; Albarqouni, S.; Wang, L.; Baust, M.; Steiger, K.; Schlitter, A. M.; Esposito, I.; and Navab, N. 2016. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Transactions on Medical Imaging*, 35(8): 1962–1971.
- Vora, J.; Bouacida, N.; Krishnan, A.; and Mohapatra, P. 2024. FedDM: Enhancing Communication Efficiency and Handling Data Heterogeneity in Federated Diffusion Models. *arXiv preprint arXiv:2407.14730*.
- Wagner, N.; Fuchs, M.; Tolkach, Y.; and Mukhopadhyay, A. 2022. Federated stain normalization for computational pathology. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 14–23. Springer.
- Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.
- Wilm, F.; Fragoso, M.; Bertram, C. A.; Stathonikos, N.; Öttl, M.; Qiu, J.; Klopffleisch, R.; Maier, A.; Breininger, K.; and Aubreville, M. 2023. Multi-scanner canine cutaneous squamous cell carcinoma histopathology dataset. In *BVM Workshop*, 206–211. Springer.
- Wu, N.; Yu, L.; Yang, X.; Cheng, K.-T.; and Yan, Z. 2023. FedIIC: Towards Robust Federated Learning for Class-Imbalanced Medical Image Classification. In Greenspan, H.; Madabhushi, A.; Mousavi, P.; Salcudean, S.; Duncan, J.; Syeda-Mahmood, T.; and Taylor, R., eds., *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, 692–702. Cham: Springer Nature Switzerland. ISBN 978-3-031-43895-0.
- Zhang, G.; Beitollahi, M.; Bie, A.; and Chen, X. 2023. Normalization is all you need: Understanding layer-normalized federated learning under extreme label shift. *arXiv preprint arXiv:2308.09565*.
- Zhang, M.; Marklund, H.; Dhawan, N.; Gupta, A.; Levine, S.; and Finn, C. 2021. Adaptive risk minimization: Learning to adapt to domain shift. *Advances in Neural Information Processing Systems*, 34: 23664–23678.
- Zhou, T.; and Konukoglu, E. 2023. Fedfa: Federated feature augmentation. *arXiv preprint arXiv:2301.12995*.
- Zhu, G.; Li, D.; Gu, H.; Yao, Y.; Fan, L.; and Han, Y. 2025. FedMIA: An Effective Membership Inference Attack Exploiting “All for One” Principle in Federated Learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 20643–20653.
- Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.
- Zhu, L.; Liu, Z.; and Han, S. 2019. Deep leakage from gradients. *Advances in neural information processing systems*, 32.
- Zingman, I.; Frayle, S.; Tankoyeu, I.; Sukhanov, S.; and Heinemann, F. 2024. A comparative evaluation of image-to-image translation methods for stain transfer in histopathology. In *Medical Imaging with Deep Learning*, 1509–1525. PMLR.

Pseudocode for FedSDA

In this section, we present two pseudocodes, including

- Target distribution fitting in Algorithm 1
- Stain alignment in Algorithm 2.

Algorithm 1: Target Distribution Fitting

Input: Number of client K , number of communication rounds R , number of local epochs E , local batch size B , learning rate η , number of diffusion timesteps T , variance schedule $\{\beta_1, \beta_2, \dots, \beta_T\}$, local histopathological image datasets $\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_K\}$, initialized $\mathcal{G}_\theta$ (parameterized by θ)

Output: Conditional diffusion model $\mathcal{G}_\theta$

```

 $|\mathcal{X}| \leftarrow \sum_{k=1}^K |\mathcal{X}_k|$ 
 $\theta_0 \leftarrow \theta$ 
# Obtaining stain matrices using stain separation
for  $k = 1$  to  $K$  do
   $\mathcal{W}_k \leftarrow \{\}$ 
  for  $\mathbf{x} \in \mathcal{X}_k$  do
     $\mathbf{w}, \_ \leftarrow \mathcal{S}(\mathbf{x})$ 
     $\mathcal{W}_k \leftarrow \mathcal{W}_k \cup \{\mathbf{w}\}$ 
  end for
end for
# Training a conditional diffusion model
for  $r = 1$  to  $R$  do
  # Local training
  for  $k = 1$  to  $K$  do
     $\theta_r^k \leftarrow \theta_{r-1}$ 
     $\mathcal{B}_k \leftarrow$  (split  $\mathcal{W}_k$  into batches of size  $B$ )
    for  $e = 1$  to  $E$  do
      for  $b \in \mathcal{B}_k$  do
         $\mathcal{L} \leftarrow 0$ 
        for  $\mathbf{w} \in b$  do
           $t \sim \text{Uniform}(\{1, 2, \dots, T\})$ 
           $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
           $\bar{\alpha}_t \leftarrow \prod_{s=1}^t (1 - \beta_s)$ 
           $\hat{\mathcal{L}} \leftarrow \|\epsilon - \mathcal{G}_{\theta_r^k}(\sqrt{\bar{\alpha}_t} \mathbf{w} + \sqrt{1 - \bar{\alpha}_t} \epsilon, t, k)\|_2^2$ 
           $\mathcal{L} \leftarrow \mathcal{L} + \hat{\mathcal{L}}$ 
        end for
         $\theta_r^k \leftarrow \theta_r^k - \eta \nabla_{\theta_r^k} \frac{\mathcal{L}}{|b|}$ 
      end for
    end for
  end for
# Central server aggregation
   $\theta_r \leftarrow \sum_{k=1}^K \frac{|\mathcal{X}_k|}{|\mathcal{X}|} \theta_r^k$ 
end for
 $\theta \leftarrow \theta_R$ 
return  $\mathcal{G}_\theta$ 

```

Pseudocode for Amplitude Normalization in HarmoFL

In this section, we present the pseudocode for amplitude normalization in HarmoFL (Jiang, Wang, and Dou 2022) (i.e., amp-norm) in Algorithm 3.

Algorithm 2: Stain Alignment

Input: Number of client K , local histopathological image datasets $\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_K\}$, conditional diffusion model $\mathcal{G}$

Output: Aligned datasets $\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2, \dots, \hat{\mathcal{X}}_K$

```

for  $k = 1$  to  $K$  do
   $n \leftarrow |\mathcal{X}_k| / K$ 
   $\{\mathcal{X}_{k,i}\}_{i=1}^K \leftarrow$  (split  $\mathcal{X}_k$  into  $K$  subsets of size  $n$ )
  for  $i = 1$  to  $K$  do
     $\hat{\mathcal{X}}_{k,i} \leftarrow \{\}$ 
    for  $\mathbf{x} \in \mathcal{X}_{k,i}$  do
       $\_, \mathbf{h} \leftarrow \mathcal{S}(\mathbf{x})$ 
       $\mathbf{w} \leftarrow \mathcal{G}(i)$ 
       $\hat{\mathbf{x}} \leftarrow \mathcal{R}(\mathbf{w}, \mathbf{h})$ 
       $\hat{\mathcal{X}}_{k,i} \leftarrow \hat{\mathcal{X}}_{k,i} \cup \{\hat{\mathbf{x}}\}$ 
    end for
  end for
   $\hat{\mathcal{X}}_k \leftarrow \bigcup_{i=1}^K \hat{\mathcal{X}}_{k,i}$ 
end for
return  $\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2, \dots, \hat{\mathcal{X}}_K$ 

```

Algorithm 3: amp-norm

Input: Number of client K , batch size B , decay factor v , local histopathological image datasets $\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_K\}$

Output: Amplitude normalized datasets $\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2, \dots, \hat{\mathcal{X}}_K$

```

# Computing the average amplitude
for  $k = 1$  to  $K$  do
   $\mathcal{A}_k \leftarrow 0$ 
   $\mathcal{B}_k \leftarrow$  (split  $\mathcal{X}_k$  into batches of size  $B$ )
  for  $b \in \mathcal{B}_k$  do
     $\hat{\mathcal{A}}_k \leftarrow 0$ 
    for  $\mathbf{x} \in b$  do
      # Let  $\mathcal{A}_{\mathbf{x}}, \mathcal{P}_{\mathbf{x}} = \mathcal{F}(\mathbf{x})$ ,
      # where  $\mathcal{F}$  is the Fourier transform,
      #  $\mathcal{A}_{\mathbf{x}}$  is the amplitude for  $\mathbf{x}$ ,
      # and  $\mathcal{P}_{\mathbf{x}}$  is the phase for  $\mathbf{x}$ .
       $\mathcal{A}, \_ \leftarrow \mathcal{F}(\mathbf{x})$ 
       $\hat{\mathcal{A}}_k \leftarrow \hat{\mathcal{A}}_k + \mathcal{A}$ 
    end for
     $\mathcal{A}_k \leftarrow (1 - v) \mathcal{A}_k + \frac{v}{B} \hat{\mathcal{A}}_k$ 
  end for
   $\bar{\mathcal{A}} \leftarrow \frac{1}{K} \sum_{k=1}^K \mathcal{A}_k$ 
# Normalizing amplitudes of histopathological images
for  $k = 1$  to  $K$  do
   $\hat{\mathcal{X}}_k \leftarrow \{\}$ 
  for  $\mathbf{x} \in \mathcal{X}_k$  do
    #  $\mathcal{F}^{-1}$  is the inverse Fourier transform.
     $\_, \mathcal{P} \leftarrow \mathcal{F}(\mathbf{x})$ 
     $\hat{\mathbf{x}} \leftarrow \mathcal{F}^{-1}(\bar{\mathcal{A}}, \mathcal{P})$ 
     $\hat{\mathcal{X}}_k \leftarrow \hat{\mathcal{X}}_k \cup \{\hat{\mathbf{x}}\}$ 
  end for
end for
return  $\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2, \dots, \hat{\mathcal{X}}_K$ 

```

Dataset Description

Mitos & Atypia 14

Mitos & Atypia 14 (i.e., MA14) contains the whole-slide images (WSIs) of hematoxylin and eosin (H&E) stained breast cancer tissue. Since the WSIs are produced from two different scanners, they are regarded as being originated from two different clients accordingly, thereby the number of clients in a federated learning environment is 2. The WSIs were split into histopathological patch images of size 256×256 at $40\times$ magnification. We provide the numbers of WSIs and histopathological patch images in each client in Table 4 and present the sample images of each client in Fig. 8.

Client	1	2
WSIs	1,136	1,136
Histopathological Patch Images	35,996	35,993

Table 4: Numbers of WSIs and histopathological patches of each client in the Mitos & Atypia 14 dataset.

CAMELYON17

CAMELYON17 (i.e., C17) contains the WSIs of H&E stained lymph node sections. Since the WSIs were collected from 5 medical centers, the number of clients in federated learning is 5. The WSIs were split into histopathological patch images of size 96×96 at $10\times$ magnification. For data splitting in tumor classification, each client follows an 80/20 ratio as the training/test sets. We provide the numbers of WSIs and histopathological patches of each client in Table 5 and present the sample images of each client in Fig. 9.

Client	1	2	3	4	5
WSIs	10	10	10	10	10
Histopathological patch Images	59,042	34,832	85,053	129,140	146,721

Table 5: Numbers of WSIs and histopathological patch images of each client in the CAMELYON17 dataset.

AGGC22

AGGC22 (i.e., A22) contains H&E stained WSIs of prostatectomy and biopsy specimens. Since each specimen is scanned by 6 scanners, the WSIs were classified into 6 clients based on which scanners they are scanned from. The WSIs were split into histopathological patch images of size 256×256 at $20\times$ magnification. For data splitting in tumor classification, each client follows a ratio of 60/40 as the training/test sets. We provide the numbers of WSIs and histopathological patch images in each client in Table 6 and present the sample images of each client in Fig. 10.

Client	1	2	3	4	5	6
WSIs	38	38	38	37	38	22
Image Patches	119,860	254,366	210,151	152,151	217,518	115,163

Table 6: Numbers of WSIs and histopathological patch images of each client in the AGGC22 dataset.

Visualization Results

Here, we first present the t-SNE (t-distributed Stochastic Neighbor Embedding) visualizations of each dataset before and after applying FedSDA, as shown in Figs. 11, 12, and 13. We observe that, after applying FedSDA, the low-dimensional features of each client’s images tend to be distributed more identically compared to those without applying FedSDA. Note that although the low-dimensional features of each client’s images in the A22 dataset are less identically distributed due to sampling bias, the results still show marginal improvement.

Second, we visualize stain matrices of each dataset, including those decomposed from histopathological patch images and those generated using the diffusion model, as shown in Figs. 14, 15, and 16. We observe that the stain matrices generated using the diffusion model for one client are almost identically distributed to those obtained from decomposing the histopathological patch images of the same client.

Finally, we further present the sample images of each client in the C17 and A22 datasets after applying FedSDA, CCST (Chen et al. 2023), and amp-norm, respectively, in Figs. 17 and 18. From the results, we can observe that CCST (Chen et al. 2023) significantly alters not only the stains of images but also their structures. In contrast, the structural alterations by amp-norm and FedSDA are not visually significant. On the other hand, across all clients, the stains in images with amp-norm applied are more monotone compared to those with FedSDA applied.

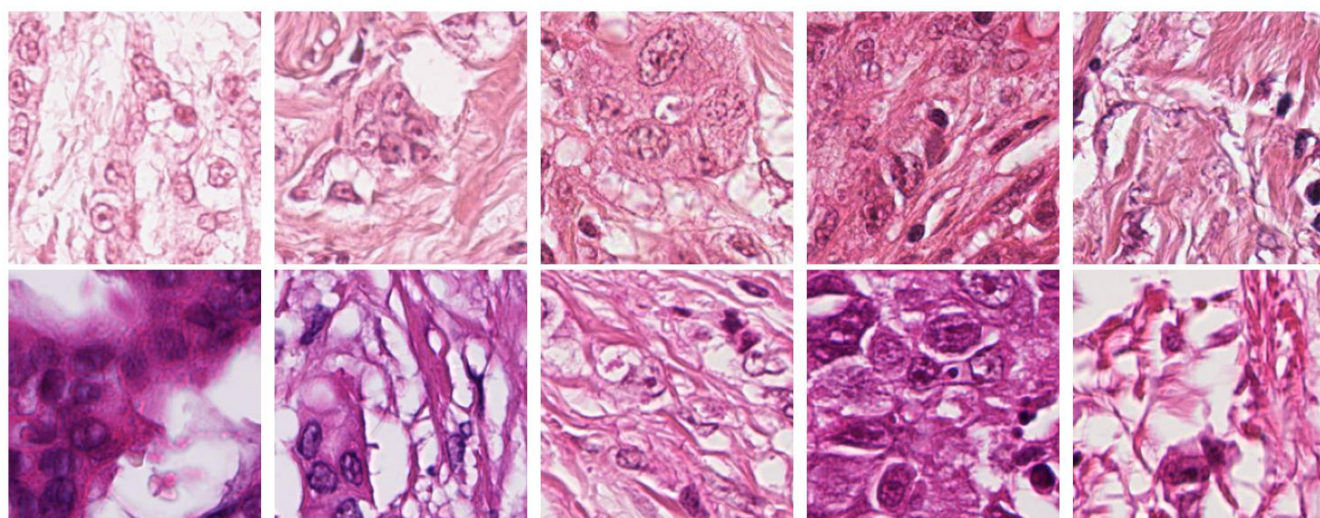


Figure 8: Sample images of the Mitos & Atypia 14 (Roux 2014) dataset. Top row: Client 1. Bottom row: Client 2.

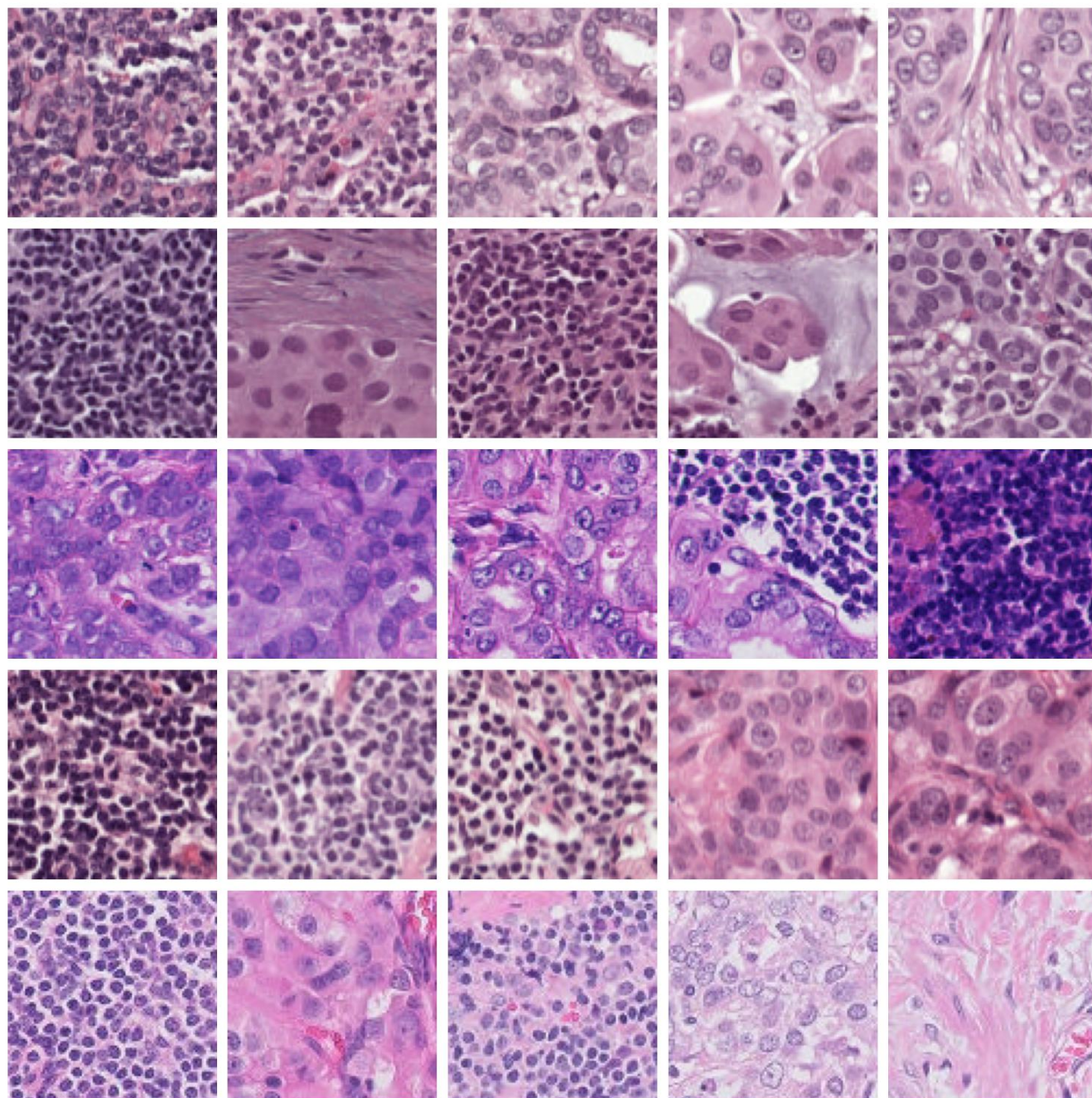


Figure 9: Sample images of the CAMELYON17 dataset (Bándi et al. 2019). Different rows, from top to bottom, indicate different clients.

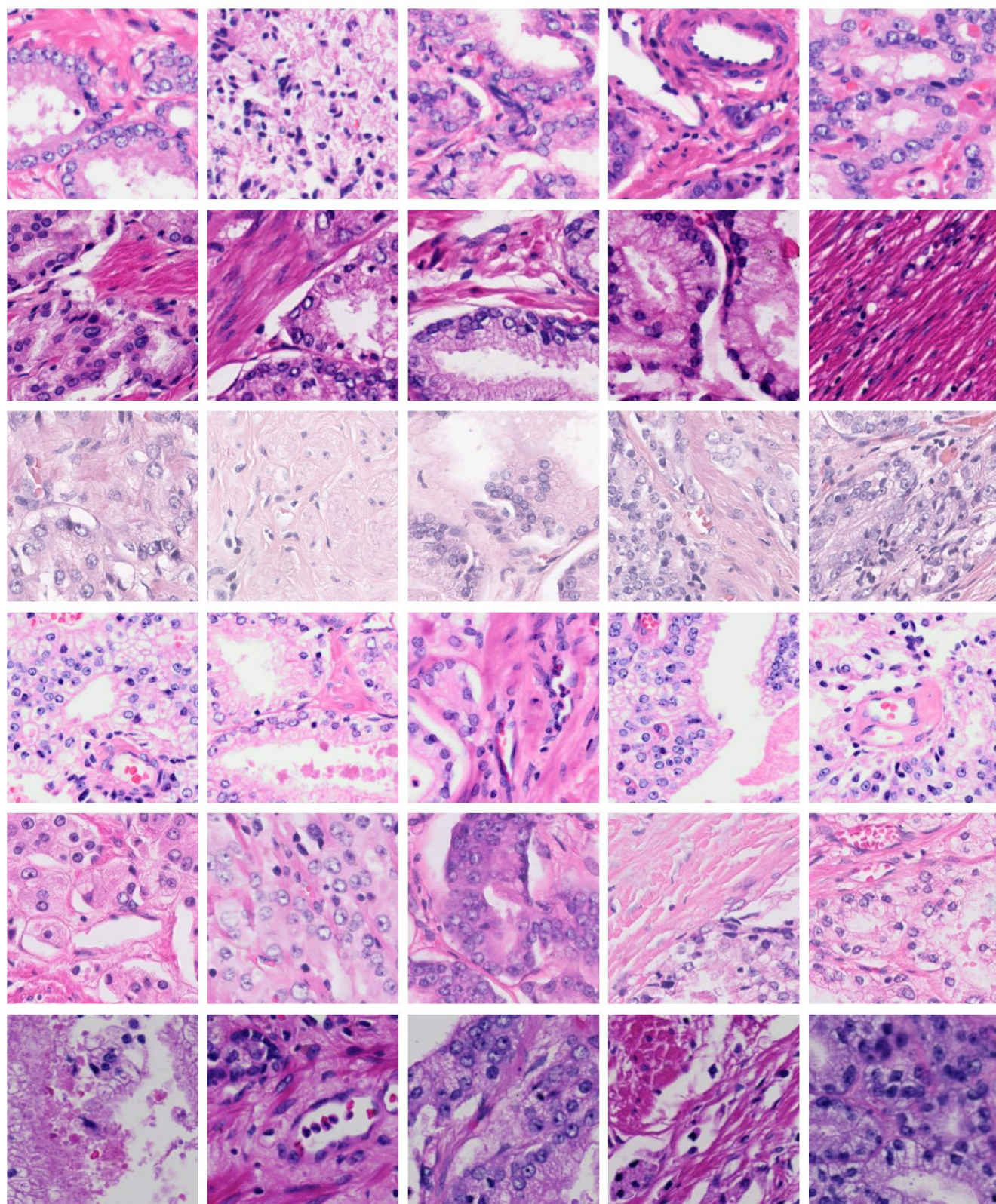


Figure 10: Sample images of the AGGC22 dataset (Huo et al. 2024). Different rows, from top to bottom, indicate different clients.

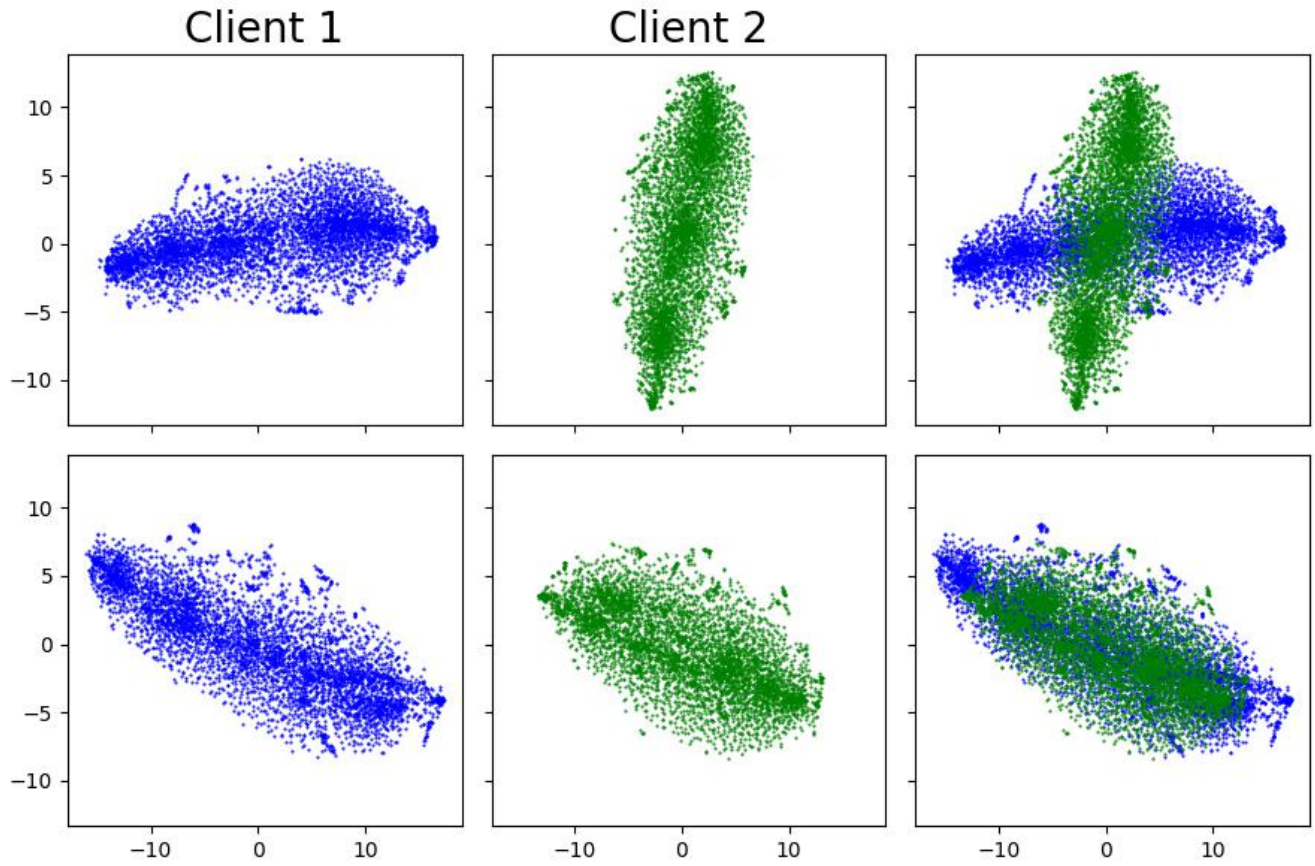


Figure 11: Visualizations of the MA14 (Roux 2014) dataset for each client before and after applying FedSDA. Top row: Before applying FedSDA. Bottom row: After applying FedSDA. The rightmost column shows the results when those from the two clients are put together.

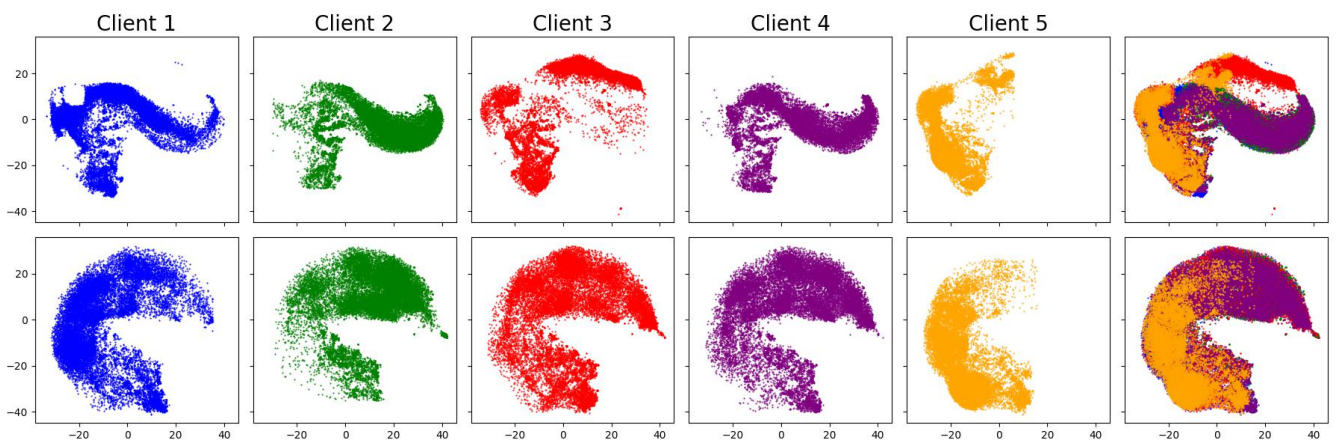


Figure 12: Visualizations of the C17 dataset (Bándi et al. 2019) before and after applying FedSDA. Top row: Before applying FedSDA. Bottom row: After applying FedSDA. The rightmost column shows the results when those from the five clients are put together.

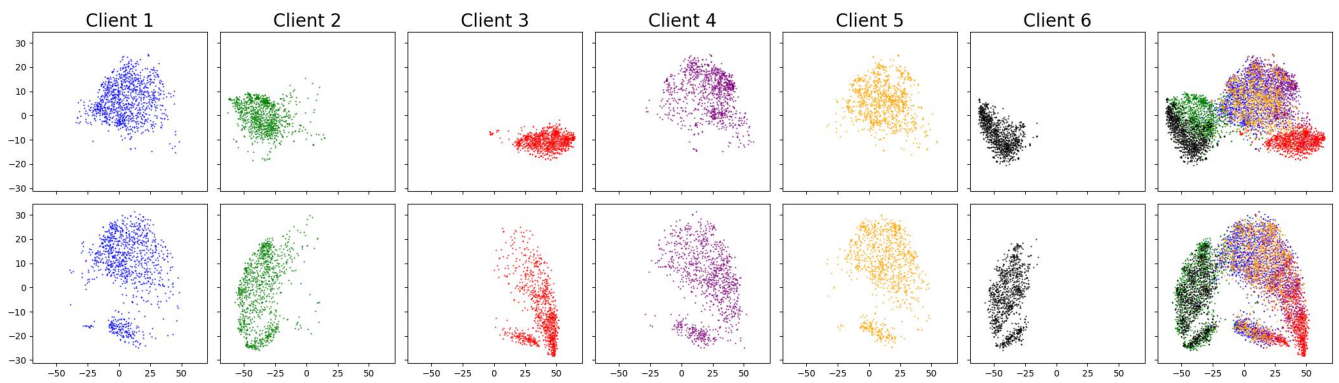


Figure 13: Visualizations of the A22 dataset (Huo et al. 2024) before and after applying FedSDA. Top row: Before applying FedSDA. Bottom row: After applying FedSDA. The rightmost column shows the results when those from the six clients are put together.

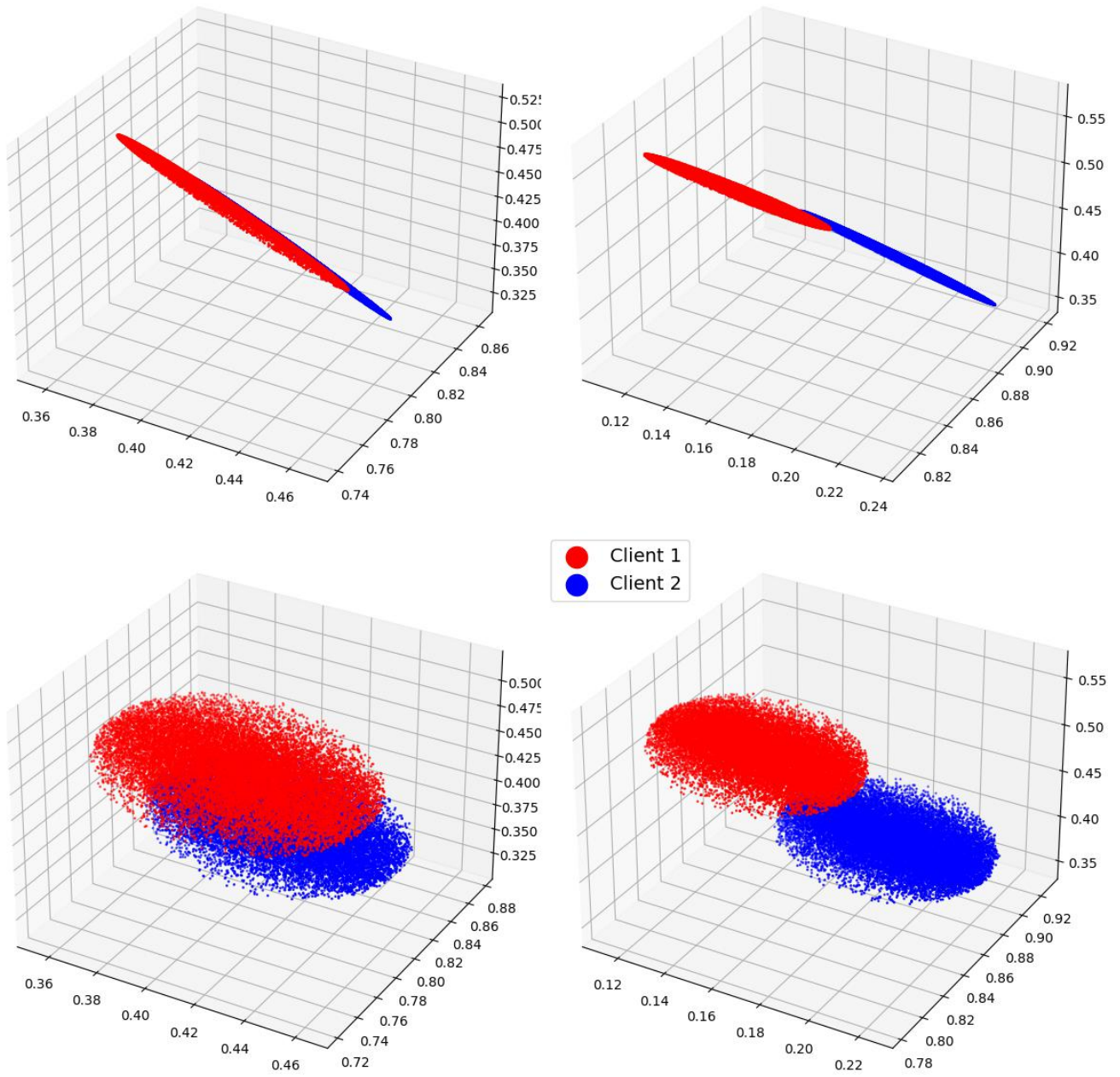


Figure 14: Stain matrix visualizations of the MitoS & Atypia 14 dataset (Roux 2014). Left column: H stain. Right column: E stain. Top row: Decomposition from histopathological images. Bottom row: Generated from a diffusion model.

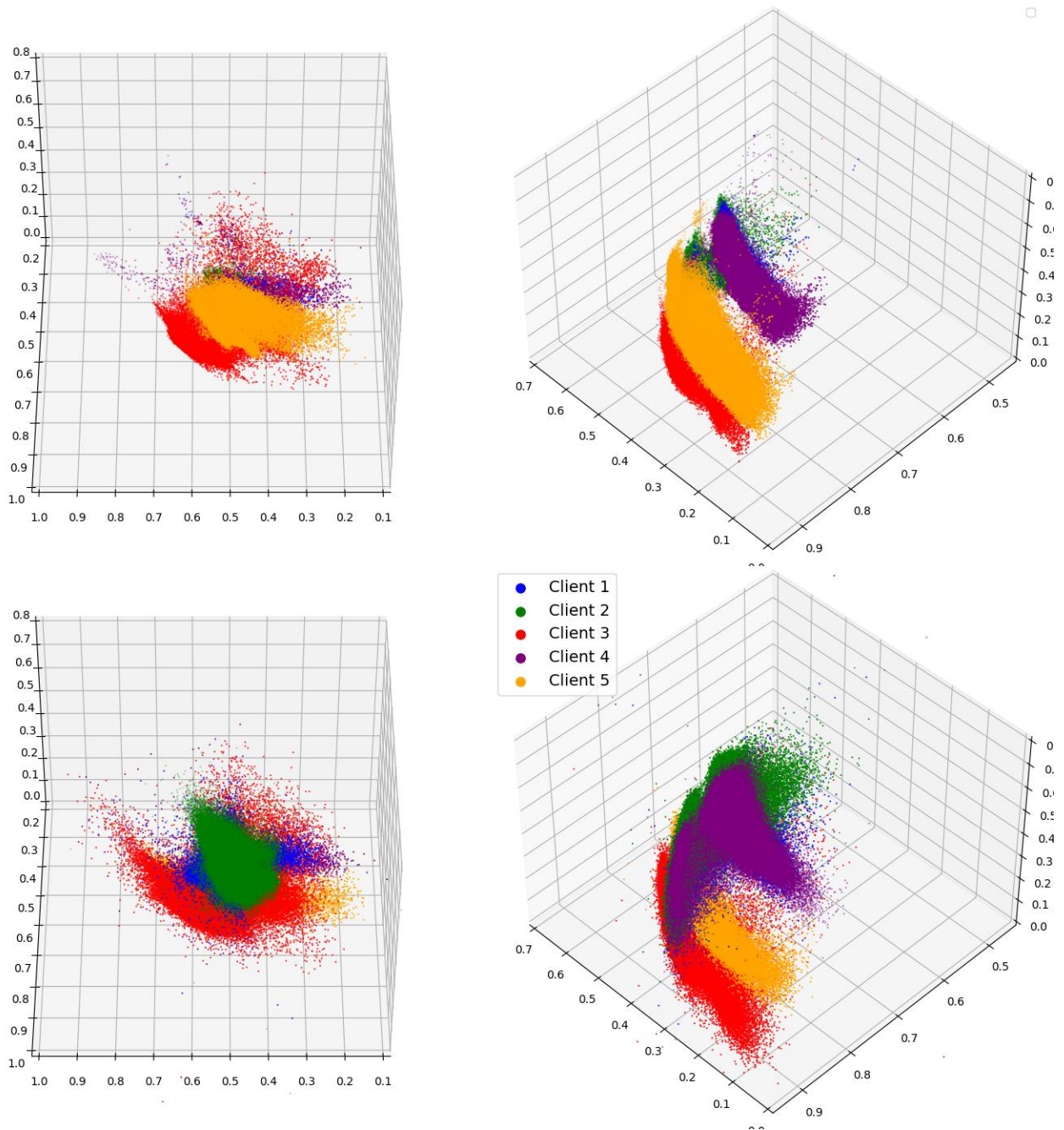


Figure 15: Stain matrix visualizations of the CAMELYON17 dataset (Bándi et al. 2019). Left column: H stain. Right column: E stain. Top row: Decomposition from histopathological images. Bottom row: Generated from a diffusion model.

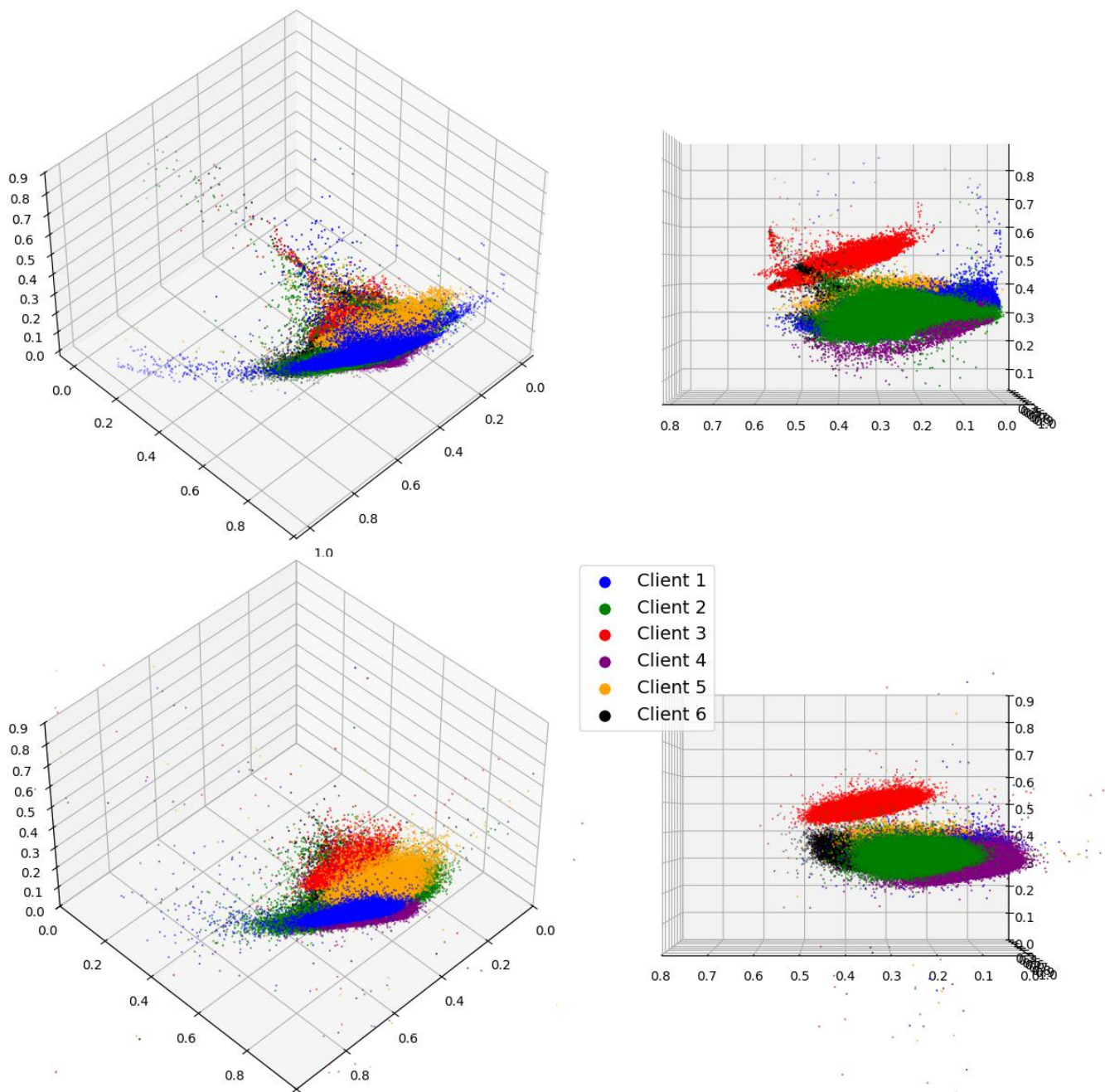


Figure 16: Stain matrix visualizations of the AGGC22 dataset (Huo et al. 2024). Left column: H stain. Right column: E stain. Top row: Decomposition from histopathological images. Bottom row: Generated from a diffusion model.

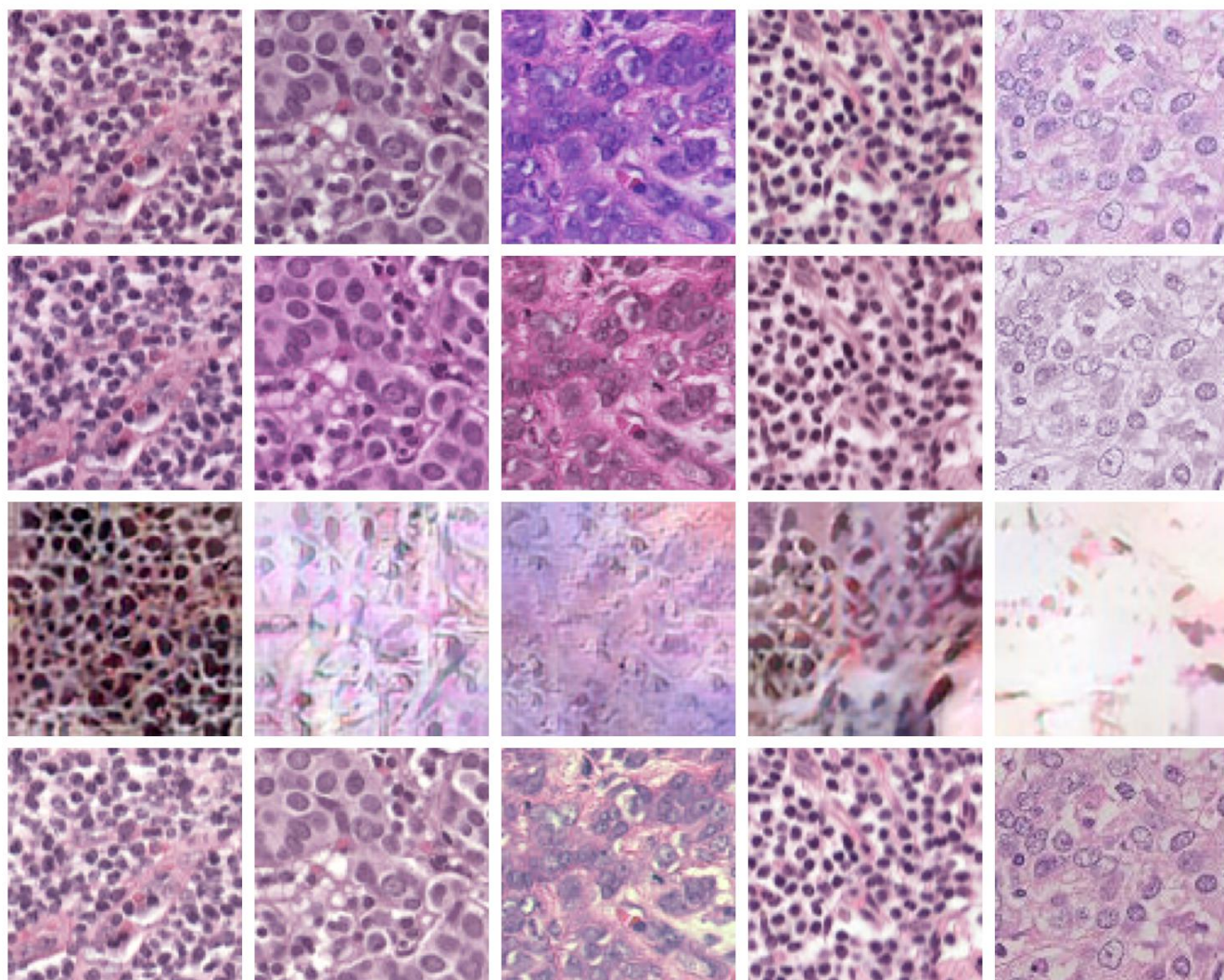


Figure 17: Sample images of the CAMELYON17 (Bándi et al. 2019) dataset, before and after applying FedSDA, CCST (Chen et al. 2023), and amp-norm. From left to right, it indicates from the 1st client to the 5th client. Top row: Original images. Second row: Images with FedSDA applied. Third row: Images with CCST (Chen et al. 2023) applied. Bottom row: Images with amp-norm applied.

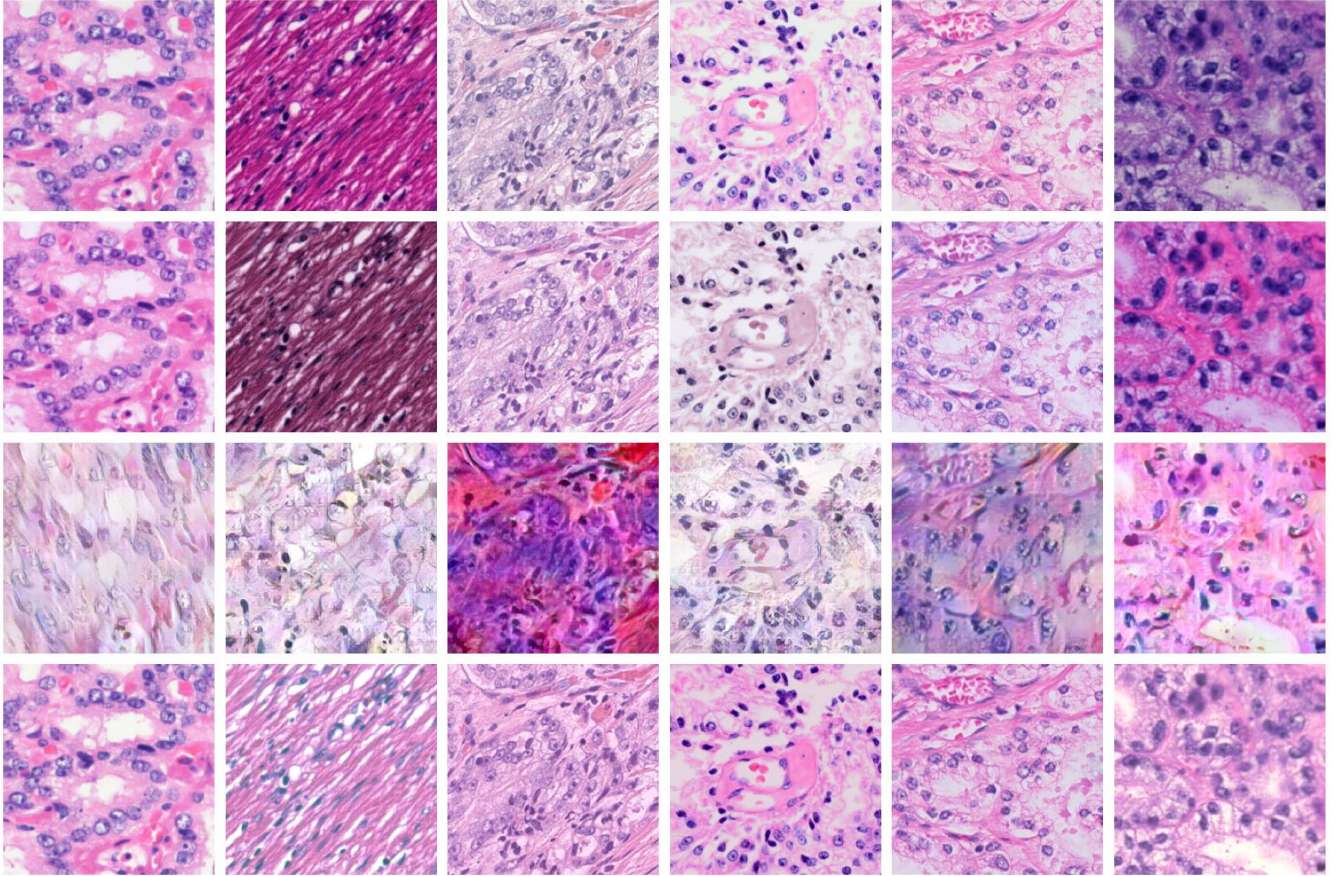


Figure 18: Sample images of the AGGC22 dataset (Huo et al. 2024), before and after applying FedSDA, CCST (Chen et al. 2023), and amp-norm. From left to right, it indicates from the 1st client to the 5th client. Top row: Original images. Second row: Images with FedSDA applied. Third row: Images with CCST (Chen et al. 2023) applied. Bottom row: Images with amp-norm applied.