pFedSolar: Personalized Federated Learning for Photovoltaic Panel Soiling Recognition

Anonymous authors
Paper under double-blind review

Abstract

The intelligent operation and maintenance (O&M) of photovoltaic (PV) power plants is crucial for ensuring their power generation efficiency and economic benefits, among which the automatic recognition of PV panel soiling types is one of the core tasks. However, subject to variations in illumination, climate, and pollution sources, PV power plants across different geographical locations exhibit significant non-independent and identically distributed (non-IID) characteristics in their soiling images. Furthermore, commercial confidentiality and location privacy concerns restrict the aggregation of individual power plant data into large-scale centralized datasets, resulting in widespread "data silos." To tackle these challenges, we propose pFedSolar, a personalized federated learning (PFL) approach tailored for PV panel soiling recognition, which leverages the vision-language foundation model CLIP. This method incorporates three core components: a federated domain adversarial learning module, a decoupled text embedding mechanism, and a personalized federated aggregation and optimization strategy. Extensive experiments on a composite dataset demonstrate that pFedSolar significantly outperforms state-of-the-art (SOTA) methods, ablation studies also confirm the efficacy of each core component.

1 Introduction

Automatic recognition of PV panel soiling types is a core task in intelligent O&M. However, PV power plants located in different geographical regions are subject to varying illumination, climatic conditions, and pollution sources, resulting in significant non-IID characteristics in their soiling images. Furthermore, operational data from individual plants often involve commercial confidentiality and geographical privacy concerns, making it difficult to aggregate them into large-scale centralized datasets (Fan et al., 2025; Miao, 2026). To overcome the dual constraints of data heterogeneity and privacy preservation, FL has been introduced into this field as a distributed learning paradigm. Concurrently, the vision-language model CLIP enables few-shot classification without extensive labeled samples, offering a novel approach to mitigate sample scarcity and long-tailed distribution issues. Integrating PFL with CLIP has emerged as a promising frontier for simultaneously addressing the triple challenges of data heterogeneity, privacy preservation, and sample scarcity.

Despite the promising prospects of this direction, existing methods still face critical limitations. Most approaches rely solely on a single learnable text embedding for federated aggregation, lacking effective mechanisms to handle scene-specific visual domain shifts across different PV power plants. Furthermore, current methods predominantly employ a uniform parameter aggregation strategy, leading to two major issues: (1) forced aggregation of scene-specific visual adaptation parameters dilutes local knowledge, thereby degrading local performance; (2) uniform text embedding aggregation causes negative transfer under highly heterogeneous scenarios.

To overcome the aforementioned limitations, this paper proposes pFedSolar, a PFL method tailored for PV panel soiling recognition. pFedSolar incorporates three core components: a federated domain adversarial learning mechanism to extract scene-invariant features, a decoupled text embedding learning framework that separates semantic knowledge into shared and private spaces, and a personalized federated aggregation strategy that differentiates component roles for selective aggregation. These three mechanisms work synergistically to achieve an optimal trade-off between global knowledge sharing and local scene adaptation.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Extensive experiments are conducted on a composite dataset, the Solar Panel Contaminated Surface Images Dataset, under two heterogeneous partitioning schemes (mild and severe). The experimental results demonstrate that pFedSolar significantly outperforms existing SOTA methods across both levels of data heterogeneity, thereby validating its effectiveness and robustness in adapting to diverse PV power plant environments.

2 Related Work

2.1 PV Panel Fault Detection and Classification

With the large-scale development of the PV industry, PV panel fault detection has become a core research direction in intelligent operation and maintenance (O&M). For example, Adib et al. (2025) proposed a deep learning framework integrated with explainable artificial intelligence (XAI) to achieve PV panel segmentation and fault classification. Yin & Gao (2024) achieves multi-plant collaborative modeling by uploading only gradient parameters, effectively safeguarding the privacy data generated by each power plant during operation; Kazemi (2025) applied federated transfer learning to PV panel fault detection, leveraging source domain knowledge to assist few-shot training in the target domain and alleviating the scarcity of cross-domain data. Systematic reviews have demonstrated that intelligent monitoring and cleaning systems are crucial for improving PV performance, yet existing methods still face severe challenges in practical deployment (Olorunfemi et al., 2022). If some defects are misclassified as other soiling types and inappropriate cleaning strategies are applied, secondary damage may even be inflicted on the modules (Mansur et al., 2026). These issues drive the exploration of distributed learning and few-shot recognition technologies in the PV O&M domain.

2.2 Vision-Language Pre-trained Models (VLM) and CLIP

Vision-language pre-trained models (VLMs) provide a novel approach to PV panel soiling recognition. As a representative VLM, CLIP establishes powerful cross-modal semantic alignment through contrastive learning on massive image-text pairs, enabling zero-shot or few-shot classification relying solely on natural language prompts. This characteristic is highly aligned with the practical demands of PV panel soiling recognition, CLIP's open-vocabulary recognition capability can effectively alleviate the performance degradation caused by long-tailed distributions, O&M personnel can also flexibly define or adjust soiling categories via natural language, reducing the reliance on data re-annotation for model iteration. However, directly applying CLIP to PV power plant scenarios still faces adaptation barriers, as its general knowledge cannot be readily transferred to industrial soiling recognition tasks characterized by high domain specificity and regional particularity, necessitating integration with distributed learning frameworks to achieve domain adaptation.

2.3 Combination of PFL and CLIP

Recent studies have explored the combination of PFL and CLIP from multiple dimensions. FedCLIP designs a lightweight adapter fine-tuning mechanism that enables rapid generalization and personalization of CLIP in federated settings by updating only a small number of parameters, significantly reducing communication overhead and computational costs (Lu et al., 2023); FedPGP preserving CLIP's general semantic knowledge through a hierarchical prompt strategy while adapting to each client's specific data distribution via local prompt vectors, thereby avoiding catastrophic forgetting caused by full fine-tuning (Cui et al., 2024); FedCoPL introduces a collaborative pseudo-labeling mechanism that allows clients to mutually refine pseudo-labels based on confidence thresholds during federated training, effectively alleviating the impact of label noise and missing categories on personalized models (Guo et al., 2025); TriplePlay combines CLIP's deep semantic understanding capability with the resource efficiency of federated learning, achieving more robust cross-domain generalization under non-IID conditions through dynamic routing and knowledge distillation strategies (Imteaj et al., 2024). These works laying a theoretical and methodological foundation for building PV panel soiling recognition technologies that are privacy-preserving, highly generalizable, and adaptable to regional differences.

3 DATASET PROCESSING

3.1 DATASET SOURCE AND CONSTRUCTION

The dataset we used was named Solar Panel Contaminated Surface Images, the main contents of it are sourced from two publicly available datasets on Kaggle. The first dataset is Solar Panel[1], the second dataset is Solar Panel Images Clean and Faulty Images[2]. The images in these datasets cover several major soiling types commonly encountered in the daily O&M of PV panels.

By integrating samples from the two aforementioned public datasets, our composite dataset comprises a total of 1,724 color real-world PV panel images. All images were captured in real-world settings and faithfully reflect the actual states of PV panel surfaces under varying environmental conditions. Representative sample images from the five soiling categories are provided in Appendix A.

3.2 DATA PARTITION STRATEGY

To reflect the reality that PV power plants across different regions encounter varying soiling types and severity levels owing to differences in geographic location and climatic conditions, we partitions the dataset into four simulated sites (four clients in PFL), each representing a typical PV power plant operating environment:

- **Site 1 — Coastal PV Power Plant:** Coastal power plants are continuously exposed to high-humidity maritime climates with frequent seabird activity, and bird-drop accumulation is one of the most common soiling forms at these facilities; concurrently, salt-spray erosion accelerates the damage and aging of PV panel surfaces (Deng & Davydov, 2023; Kshetrimayum et al., 2023).
- **Site 2 — Desert PV Power Plant:** PV power plants in arid desert regions are persistently subjected to intense sunlight and wind-driven sand erosion, and the continuous deposition of dust on PV panel surfaces makes dust accumulation particularly acute, severely affecting photoelectric conversion efficiency (Shi et al., 2025; Ezzini et al., 2025; Turag et al., 2022).
- **Site 3 — Cold PV Power Plant:** Cold regions are characterized by long winters with frequent snowfall accompanied by subfreezing temperatures, and extensive snow coverage on PV panel surfaces constitutes the primary O&M challenge encountered by such plants during winter (Rezzouk et al., 2025).
- **Site 4 — Temperate PV Power Plant:** Large ground-mounted PV power plants located in temperate climate zones experience relatively mild climatic conditions, with diverse soiling types and no single dominant category; the distribution across soiling types is relatively balanced.

The selection of the aforementioned four simulated sites is not arbitrary but rather grounded in the typical geographic distribution of PV power plants worldwide. Coastal areas, arid deserts, cold regions, and temperate plains all host large-scale centralized PV power plants, whose climatic conditions and soiling factors differ significantly (Gunes & Balog, 2025; Boyle et al., 2017). Partitioning the dataset according to these four categories can more faithfully reflects the distinct O&M challenges encountered by power plants across different regions, thereby enhancing the practical engineering significance of the experimental conclusions.

Regarding soiling category classification, we categorizes PV panel conditions into five categories: Bird-drop, Clean, Damage (physical/electrical), Dusty, and Snow-Covered. Notably, physical damage and electrical damage are consolidated into a single category because both exhibit similar visual features characterized by structural damage to PV panel surfaces and require similar repair strategies in practical O&M.

3.3 HETEROGENEITY SETTINGS

To comprehensively evaluate the performance of the proposed method under varying degrees of data heterogeneity, this paper develops two data partitioning schemes to simulate mild and severe

[1] https://www.kaggle.com/datasets/rppost/solar-panel

[2] https://www.kaggle.com/datasets/pythonafroz/solar-panel-images

heterogeneity scenarios, respectively. The mild scheme models realistic class imbalances across climatic regions, while the severe scheme represents an extreme setting where each site is dominated by a single soiling type. Detailed descriptions of both schemes and the full sample allocation counts are provided in Appendix A.

4 METHOD

This section provides a formal description of the proposed PV panel soiling recognition method, which integrates PFL with CLIP. Section 4.1 reviews the foundational principles of CLIP and the problem formulation. Section 4.2 outlines the overall framework and highlights the key challenges. Subsequently, we detail the proposed method, which is built upon three core components: the federated domain adversarial learning strategy (Section 4.3), the decoupled text embedding learning mechanism (Section 4.4), and the personalized federated aggregation and optimization strategy (Section 4.5).

4.1 PRELIMINARIES

We adopt CLIP (Radford et al., 2021) as the foundation model for our framework. CLIP features a dual-branch architecture comprising an image encoder and a text encoder, which compute classification probabilities via cosine similarity between visual and textual embeddings scaled by a temperature parameter τ.

In the PFL-based PV panel soiling recognition setting, we consider K clients corresponding to distinct PV plant scenarios, each holding a local dataset drawn from a scenario-specific distribution. Significant domain discrepancies exist across plants due to varying illumination, climate, and imaging backgrounds. Our objective is to collaboratively learn a soiling recognition model across all clients without sharing raw image data, balancing cross-plant generalization with scenario-specific adaptation.

4.2 FRAMEWORK OVERVIEW

To tackle the challenges of cross-plant domain shifts and data heterogeneity in PV panel soiling recognition, we propose a novel PFL framework that integrates three complementary strategies. The overall architecture is illustrated in Figure 1.

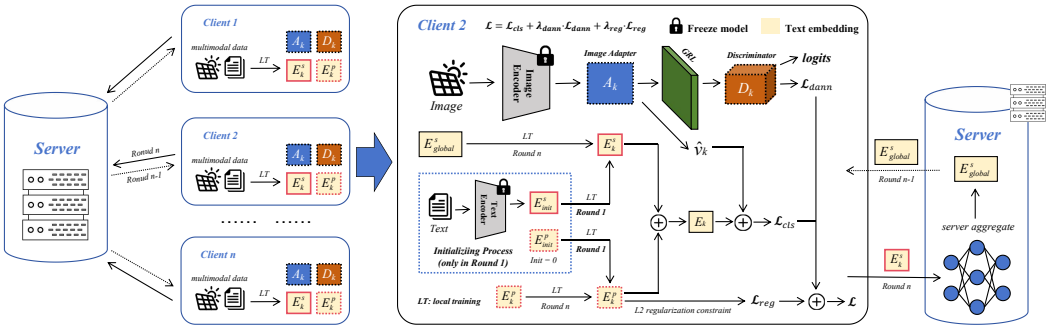

Figure 1: Overall framework of pFedSolar. during each round of local training at a client, the local image adapter A_k, domain discriminator D_k, shared embeddings $\mathbf{E}_k^s$, and private embeddings $\mathbf{E}_k^p$ are updated. dashed outlines denote A_k, D_k, and $\mathbf{E}_k^p$, indicating that they are excluded from the federated aggregation process. conversely, the solid outline of $\mathbf{E}_k^s$ signifies its participation in federated communication.

Specifically, each client k is initialized with the same pre-trained CLIP model. The backbone parameters are kept entirely frozen to preserve rich pre-trained knowledge and prevent catastrophic forgetting. Built upon the frozen CLIP, each client maintains three lightweight learnable modules: (1) an image adapter A_k, which adapts visual features to the specific plant scenario; (2) a set of

decoupled text embeddings, comprising shared embeddings $\mathbf{E}_k^s$ and private embeddings $\mathbf{E}_k^p$; and (3) a local domain discriminator D_k for scene adversarial training.

The overall objective function is formulated as:

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda_{\text{dann}} \cdot \mathcal{L}_{\text{dann}} + \lambda_{\text{reg}} \cdot \mathcal{L}_{\text{reg}}, \tag{1}$$

where $\mathcal{L}_{\text{cls}}$ is the classification loss, $\mathcal{L}_{\text{dann}}$ is the scene adversarial loss, and $\mathcal{L}_{\text{reg}}$ is the L2 regularization term applied to the private embeddings. The hyperparameters λ_{dann} and λ_{reg} control the trade-off among these terms.

To address the challenges of cross-plant domain shifts and semantic heterogeneity, we introduce three complementary strategies: federated domain adversarial learning (Section 4.3), decoupled text embedding learning (Section 4.4), and personalized federated aggregation and optimization (Section 4.5). Together, these strategies achieve an optimal balance between global knowledge sharing and local scene adaptation.

We also incorporate domain-specific prompt engineering, with details provided in Appendix B.

4.3 Federated Domain Adversarial Learning

To learn scene-invariant features that generalize across diverse PV power plant scenarios, we introduce a federated domain adversarial learning mechanism inspired by the Domain-Adversarial Neural Network (DANN) (Ganin et al., 2016). The core idea is to train feature representations that remain indistinguishable across different scenes via a Gradient Reversal Layer (GRL). This enables the model to extract essential soiling features that are invariant to specific power plant scenarios.

Image Adapter. Inspired by FedCLIP, we employ a lightweight attention-based image adapter A_k for each client k to transform frozen CLIP visual features for scene-specific adaptation. The adapter produces adapted features through element-wise multiplication:

$$\hat{v}_k(x) = v(x) \odot A_k(v(x)), \tag{2}$$

where $\odot$ denotes element-wise multiplication, and $A_k(v(x))$ represents the attention weight vector output by the adapter.

Scene Discriminator and Gradient Reversal. Following the image adapter, we connect a scene discriminator D_k to the adapted features via a Gradient Reversal Layer (GRL). During forward propagation, the GRL acts as an identity mapping, whereas during backpropagation, it reverses the gradient direction and scales it by a factor α. The scene discriminator receives the adapted features and predicts the power plant scenario identity:

$$\hat{d}_k = D_k(R_\alpha(\hat{v}_k(x))), \tag{3}$$

$R_\alpha(\cdot)$ denote the GRL operation. where D_k outputs a K-dimensional logit vector for classifying K PV power plant scenarios. The scene adversarial loss uses cross-entropy between predicted scene labels and true identities, extending the binary DANN formulation to multi-class across K clients:

$$\mathcal{L}_{\text{dann}} = -\frac{1}{|\mathcal{B}|} \sum_{x \in \mathcal{B}} \log \frac{\exp([\hat{d}_k(x)]_k)}{\sum_{j=1}^{K} \exp([\hat{d}_k(x)]_j)}, \tag{4}$$

where $[\hat{d}_k(x)]_k$ denotes the logit corresponding to the true scenario k and $\mathcal{B}$ represents the mini-batch.

Through the GRL mechanism, the image adapter learns soiling-essential representations invariant to plant scenarios, while the discriminator optimizes to identify sample sources. This adversarial process extracts scene-agnostic adapted features, enabling the model to focus on intrinsic soiling characteristics rather than background variations. To stabilize training, we employ a linear progressive scheduling strategy for α (details in Appendix B). Each client maintains its own independent scene discriminator (not aggregated), as summarized in Table 1.

4.4 Decoupled Text Embedding Learning

Although domain adversarial learning addresses the visual scene shift problem, we observe that the text semantics in PV panel soiling recognition exhibit a dual nature. Certain soiling category semantics are universally shared across all plant scenarios, while others vary due to scenario-specific data

characteristics. To capture this duality, we propose a decoupled text embedding learning mechanism that explicitly separates text embeddings into a shared semantic space and a private semantic space.

Decoupled Embedding Formalization. For each client k, we decompose the text embeddings into two components:

$$\mathbf{E}_k = \mathbf{E}_k^s + \mathbf{E}_k^p, \tag{5}$$

where $\mathbf{E}_k^s$ denotes the shared embedding encoding universal soiling semantics across plants, and $\mathbf{E}_k^p$ captures the private embedding for scenario-specific semantic variations.

Differentiated Aggregation Strategy. The two embedding components follow distinct aggregation strategies. The shared embeddings are aggregated across clients via Federated Averaging (FedAvg) (McMahan et al., 2017):

$$\mathbf{E}_{\text{global}}^s = \sum_{k \in \mathcal{K}} \frac{n_k}{\sum_{j \in \mathcal{K}} n_j} \cdot \mathbf{E}_k^s, \tag{6}$$

where n_k is the number of samples on client k. The private embeddings are strictly maintained locally and excluded from federated communication, preserving each plant's scenario-specific semantic adaptations.

L2 Regularization on Private Embeddings. To prevent private embeddings from deviating excessively from the shared semantic space, we impose an L2 regularization constraint:

$$\mathcal{L}_{\text{reg}} = \frac{1}{C} \sum_{i=1}^{C} \|\mathbf{E}_k^p[i]\|_2, \tag{7}$$

which encourages the private embeddings to capture only mild scenario-specific adjustments.

4.5 Personalized Federated Aggregation and Optimization Strategy

Beyond the aforementioned core components, we further design a personalized federated aggregation and optimization strategy. This strategy applies tailored processing to each model component based on its specific role within the framework, serving as a cornerstone for achieving the objectives of PFL.

Federated Aggregation Strategy. We adopt a fine-grained aggregation strategy that applies tailored processing to distinct model components during federated communication, as summarized in Table 1.

Table 1: Aggregation policy

Component	Aggregation	Design Motivation
CLIP Backbone	Frozen	Preserve pre-trained knowledge and prevent catastrophic forgetting
Image Adapter A_k	Local only	Retain scene-specific visual adaptations for each plant
Shared Embedding $\mathbf{E}_k^s$	FedAvg	Share universal soiling semantics across plants
Private Embedding $\mathbf{E}_k^p$	Local only	Capture scenario-specific semantic variations locally
Scene Discriminator D_k	Local only	Maintain local discriminative capability for each plant

As summarized in Table 1, only the shared embeddings are aggregated across clients via FedAvg. The CLIP backbone remains frozen to preserve pre-trained knowledge, while the image adapter, private embeddings, and scene discriminator are maintained locally for each client. This selective aggregation strategy departs from traditional FL methods that uniformly aggregate all parameters, striking an optimal balance between global generalization and local adaptation.

Optimization. During local training, each client minimizes a combined objective comprising classification loss, scene adversarial loss, and regularization loss:

$$\mathcal{L}_k = \mathcal{L}_{\text{cls}}(\hat{v}_k(x), \mathbf{E}_k) + \lambda_{\text{dann}} \cdot \mathcal{L}_{\text{dann}}(\hat{v}_k(x)) + \lambda_{\text{reg}} \cdot \mathcal{L}_{\text{reg}}(\mathbf{E}_k^p), \tag{8}$$

where $\mathcal{L}_{\text{cls}}$ is the cross-entropy loss, and $\hat{v}_k(x) = v(x) \odot A_k(v(x))$ denotes the scene-adapted visual features. Classification is performed via cosine similarity between the adapted visual features and the effective text embeddings $\mathbf{E}_k = \mathbf{E}_k^s + \mathbf{E}_k^p$. Only learnable parameters are updated; the CLIP backbone remains frozen.

The detailed communication protocol and industrial deployment perspective analysis are provided in Appendix B.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP

Datasets. We evaluate the proposed method on our composite dataset and two partitioning schemes with different heterogeneous degrees described in Section 3.

Comparison Methods. To comprehensively validate the performance of the proposed method, we compare it against traditional FL methods and recent state-of-the-art (SOTA) PFL approaches.

- **FedAvg**: The classic federated averaging algorithm that aggregates model parameters across all clients via weighted averaging, serving as the most fundamental baseline in FL.
- **FedProx** (Li et al., 2020): Extends FedAvg by introducing a proximal term constraint, which restricts the deviation of local model updates from the global model to mitigate performance degradation caused by data heterogeneity.
- **FedPGP**: A personalized federated prompt learning method based on privacy-preserving global personalization. It balances generalization and personalization through a hierarchical prompt strategy, retaining CLIP's general semantic knowledge while adapting to local data distributions via local prompt vectors.
- **FedCoPL**: A cooperative pseudo-labeling method designed for unsupervised federated classification. Clients estimate and upload pseudo-label distributions, which the server adjusts and redistributes to alleviate global class imbalance. It employs a partial prompt aggregation strategy that aggregates only visual prompts while keeping text prompts local.
- **FedMGP** (Bo et al., 2025): Achieves PFL through the collaborative learning of multiple text-visual prompt groups. Each client maintains multiple paired prompt groups, ensuring that each group captures distinct semantic features via a diversity loss. Furthermore, it adopts a similarity-based dynamic prompt selection strategy for aggregation.
- **FAA-CLIP** (Wu et al., 2025): A federated adversarial adaptation method that adapts the CLIP model using a lightweight Feature Adaptation Module (FAM) and a Domain Adversarial (DA) module. Specifically, FAM adapts CLIP features via attention masks, while DA learns domain-invariant representations through a domain discriminator.

Implementation Details. Detailed experimental settings, including model backbone, training configurations, and hyperparameters, are provided in Appendix C.

5.2 ABLATION STUDY

To verify the effectiveness of each core component in the proposed method, we conduct a stepwise ablation study by progressively adding each novel module to a base configuration. The four variants are defined as follows:

- **v1 (Base)**: We adopt FedCLIP as the foundational baseline, which utilizes a frozen CLIP backbone and a single learnable text embedding. FedCLIP is an adapter-based method designed for rapid generalization and personalization in federated settings via lightweight adapter fine-tuning. This baseline serves to verify the incremental contribution of each subsequent module introduced in our framework.
- **v2**: Builds upon v1 by incorporating federated domain adversarial learning.
- **v3**: Extends v2 by introducing decoupled text embedding learning (comprising shared and private embeddings).
- **v4 (Full Method)**: Extends v3 by integrating the personalized federated aggregation and optimization strategy (including a frozen CLIP backbone, non-aggregated image adapter, non-aggregated private embedding, and L2 regularization constraint).

Table 2 presents the ablation study results (averaged over five seeds; per-seed details are provided in Appendix D).

Table 2: Ablation study results (average accuracy over 5 seeds). v1: base; v2: +federated domain adversarial learning; v3: +decoupled text embedding learning; v4: +personalized federated aggregation and optimization strategy (complete method). **bold** indicates the best result.

	v1	v2	v3	v4
Dataset 1 (Mild)	68.43±2.14	69.40±2.18	87.21±0.68	**92.07±0.89**
Dataset 2 (Severe)	60.36±3.77	60.72±2.87	80.87±2.50	**93.62±0.99**

As summarized in Table 2, several key observations can be drawn:

(1) Significant Contribution of Decoupled Text Embedding Learning. Comparing v2 and v3, the incorporation of decoupled text embedding learning substantially boosts the accuracy on Dataset 1 from 69.40% to 87.21% (+17.81%), and on Dataset 2 from 60.72% to 80.87% (+20.15%). This remarkable improvement demonstrates that explicitly separating text embeddings into a shared semantic space and a scene-private space, along with adopting a differentiated aggregation strategy, effectively mitigates the semantic discrepancies among PV power plants and serves as a key mechanism for balancing global knowledge sharing with local scene adaptation. The modest gain from v1 to v2 via federated domain adversarial learning provides a stable visual feature foundation for this subsequent improvement.

(2) Further Improvement from Personalized Aggregation and Optimization Strategy. Comparing v3 and v4, the introduction of the personalized federated aggregation and optimization strategy further improves the accuracy on Dataset 1 from 87.21% to 92.07% (+4.86%), and substantially boosts the accuracy on Dataset 2 from 80.87% to 93.62% (+12.75%). Notably, this strategy yields a more significant gain on Dataset 2, which exhibits a higher degree of heterogeneity. This indicates that the component-level differentiated aggregation, the L2 regularization constraint, and the frozen CLIP backbone effectively prevent overfitting and enhance the model's generalization capability in highly heterogeneous scenarios.

(3) Impact of the Degree of Heterogeneity. Comparing the two datasets, the accuracy of v1, v2 and v3 on Dataset 2 is lower than that on Dataset 1 due to the higher degree of heterogeneity. However, the proposed complete method (v4) achieves a high accuracy of 93.62% on Dataset 2, even surpassing the 92.07% obtained on Dataset 1. This demonstrates that the proposed method possesses superior robustness in handling extreme data heterogeneity, which holds significant practical value for deployment in diverse PV power plant environments. In practical scenarios, the data distribution discrepancies among plants across different geographical locations are often far more extreme than "mild heterogeneity". The fact that pFedSolar's performance not only avoids degradation but actually improves under extreme heterogeneity, while maintaining excellent accuracy without per-site parameter tuning, significantly reduces the engineering complexity of large-scale deployment.

5.3 Comparative Experiments

Tables 3 and 4 present the comparative results of the proposed method against existing FL baselines and SOTA PFL approaches on Dataset 1 (mild heterogeneity) and Dataset 2 (severe heterogeneity), respectively.

Table 3: Comparison results with existing methods on Dataset 1 (mild heterogeneity). accuracy (%) is reported. **bold** indicates the best result.

	FedAvg	FedProx	FedPGP	FedCoPL	FedMGP	FAA-CLIP	pFedSolar
Seed 0	28.20	31.02	86.18	56.93	83.66	91.63	**93.04**
Seed 42	23.50	32.84	83.89	53.34	81.80	**93.37**	93.04
Seed 123	34.17	33.71	83.11	59.07	82.84	89.58	**91.36**
Seed 456	36.32	38.65	82.93	58.90	84.10	89.39	**91.63**
Seed 789	30.21	33.10	83.35	54.65	82.49	89.35	**91.30**
Avg	30.48±5.04	33.86±2.86	83.89±1.33	56.58±2.55	82.98±0.92	90.66±1.79	**92.07±0.89**

Table 4: Comparison results with existing methods on Dataset 2 (severe heterogeneity). accuracy (%) is reported. **bold** indicates the best result.

	FedAvg	FedProx	FedPGP	FedCoPL	FedMGP	FAA-CLIP	pFedSolar
Seed 0	38.01	33.68	77.33	52.56	83.86	86.33	**92.86**
Seed 42	26.79	20.15	77.35	58.35	84.62	88.18	**94.90**
Seed 123	33.42	30.61	80.13	57.12	85.01	83.82	**92.60**
Seed 456	30.87	28.32	80.10	50.04	84.87	85.74	**94.39**
Seed 789	28.06	20.92	83.17	56.60	84.86	82.37	**93.37**
Avg	31.43±4.48	26.74±5.98	79.62±2.42	54.93±3.49	84.64±0.46	85.29±2.25	**93.62±0.99**

As summarized in Tables 3 and 4, the following key observations can be drawn:

(1) Significant Advantage over SOTA Methods. pFedSolar significantly outperforms all SOTA PFL methods on both datasets. On Dataset 1, pFedSolar (92.07%) surpasses the best-performing SOTA method, FAA-CLIP (90.66%), by 1.41 percentage points, and exceeds FedPGP (83.89%) and FedMGP (82.98%) by 8.18 and 9.09 percentage points, respectively. On Dataset 2, pFedSolar (93.62%) outperforms FAA-CLIP (85.29%), FedMGP (84.64%), and FedPGP (79.62%) by substantial margins of 8.33, 8.98, and 14.00 percentage points, respectively. These results comprehensively demonstrate that, through the synergistic integration of federated domain adversarial learning, decoupled text embeddings, and the personalized aggregation strategy, the proposed method achieves significant performance breakthroughs over existing SOTA methods in PV panel soiling recognition.

(2) Comparison with Traditional Baselines and Unsupervised method. pFedSolar's advantage over traditional FL baselines (FedAvg, FedProx) is even more pronounced, with both methods achieving only 20%–40% accuracy and demonstrating that unified model aggregation fundamentally fails under scenario heterogeneity. FedCoPL, an unsupervised pseudo-labeling method, also underperforms (56.58% and 54.93%) since its pseudo-labeling mechanism cannot fully leverage supervised label information. Compared to FedAvg, pFedSolar improves by 61.59 and 62.19 percentage points on the two datasets, respectively.

(3) Impact of Data Heterogeneity on SOTA Methods. Compared to Dataset 1, the performance of pFedSolar on Dataset 2 improves from 92.07% to 93.62% (+1.55%), whereas all other methods exhibit a downward trend. This further confirms the exceptional adaptability of pFedSolar in handling extreme data heterogeneity.

Additional analyses, including generalization capability, hyperparameter sensitivity, and confusion matrix analysis, are provided in Appendix E.

6 Conclusion

We propose pFedSolar, a personalized federated learning framework integrated with CLIP for PV panel soiling recognition. Through three synergistic mechanisms, pFedSolar achieves an optimal trade-off between global knowledge sharing and local scene adaptation. Extensive experiments demonstrate that pFedSolar significantly outperforms existing SOTA methods across datasets with varying degrees of heterogeneity, exhibiting superior generalization capability and robustness in extreme scenarios. In the future, we will explore theoretical analysis of pFedSolar, including convergence proofs and privacy quantification, as well as its applications to other PV module defect detection tasks and lightweight deployment on resource-constrained edge devices.

AI use statement

In this work, we used generative AI tools for language polishing and proofreading of the manuscript to improve clarity and readability. We have not used generative AI tools for generating research ideas, designing experiments, or analyzing data. We have reviewed all AI-assisted text to ensure accuracy and originality. We take responsibility for the final content of this work, including text produced with the aid of generative AI.

Ethics statement

This work uses publicly available datasets and does not involve human subjects or sensitive data. The proposed method is designed for PV panel soiling recognition, which has positive implications for renewable energy maintenance and efficiency. We are not aware of any ethical concerns related to this work.

Reproducibility statement

We provide detailed descriptions of the dataset construction, model architecture, and experimental settings in the main paper. The implementation details, including hyperparameters and training configurations, are described in Appendix C. All experiments use the publicly available CLIP ViT-B/32 backbone. We plan to release the source code and dataset upon acceptance to facilitate reproducibility.

References

Asif Ur Rahman Adib, Mainul Islam, Md. Shadman Abid, and Razzaqul Ahshan. A deep learning based framework for solar panel segmentation and fault classification enhanced with explainable ai. *Solar Energy*, 302:114058, 2025.

Weihao Bo, Yanpeng Sun, Yu Wang, Xinyu Zhang, and Zechao Li. FedMGP: Personalized federated learning with multi-group text-visual prompts. In *Advances in Neural Information Processing Systems*, pp. 29598–29613, 2025.

L. Boyle, P. D. Burton, V. Danner, M. P. Hannigan, and B. King. Regional and national scale spatial variability of photovoltaic cover plate soiling and subsequent solar transmission losses. *IEEE Journal of Photovoltaics*, 7(5):1354–1361, 2017.

Tianyu Cui, Hongxia Li, Jingya Wang, and Ye Shi. FedPGP: Harmonizing generalization and personalization in federated prompt learning. In *International Conference on Machine Learning*, pp. 9646–9661, 2024.

Yuanbiao Deng and Vadim Davydov. Search for new solutions to improve the efficiency of solar cells using in saint petersburg and hong kong. In *2023 International Conference on Electrical Engineering and Photonics (EESC)*, 2023.

Mustapha Ezzini, Mohammed Ennejjar, Raja Mouachi, and Abdelali El Gourari. Impact of dust accumulation on photovoltaic panel performance in semi-arid climates. In *2025 International Conference on Decision Aid Sciences and Applications (DASA)*, pp. 2443–2447, 2025.

Yukun Fan, Ping an Zhong, Feilin Zhu, Ran Mo, Han Wang, Jia Wei, Yurou Zeng, Bin Wang, and Xinyuan Qian. Assessing the potential and complementary characteristics of china's solar and wind energy under climate change. *Renewable Energy*, 249:123213, 2025.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016.

Ibrahim Gunes and Robert S. Balog. Examination of pollution effect for pv panels in turkey's climate conditions. In *2025 IEEE Texas Power and Energy Conference (TPEC)*, 2025.

Kuangpu Guo, Lijun Sheng, Yongcan Yu, Jian Liang, Zilei Wang, and Ran He. FedCoPL: Cooperative pseudo labeling for unsupervised federated classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3326–3336, 2025.

Ahmed Imteaj, Md. Zarif Hossain, Saika Zaman, and Abdur R. Shahid. TriplePlay: Enhancing federated learning with CLIP for non-IID data and resource efficiency. In *Proceedings of the 23rd International Conference on Machine Learning and Applications*, pp. 1474–1480, 2024.

Keyvan Kazemi. Federated transfer learning for image-based solar panel fault detection. In *2025 12th Iranian Conference on Renewable Energies and Distributed Generation (ICREDG)*, 2025.

Satchidanand Kshetrimayum, Chia-Yuan Liu, and Yo-Ping Huang. An intelligent system for automatic detection and cleaning of bird droppings on solar panels. In *2023 International Automatic Control Conference (CACS)*, 2023.

Tian Li, Ameet Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In *Proceedings of Machine Learning and Systems*, 2020.

Wang Lu, Xixu Hu, Jindong Wang, and Xing Xie. FedCLIP: Fast generalization and personalization for CLIP in federated learning. *IEEE Data Engineering Bulletin*, 46(1):52–66, 2023.

Ahmed Al Mansur, Md. Sabbir Alam, Shahariar Ahmed Himo, Khawza Iftekhar Uddin Ahmed, and Md. Fayyaz Khan. Impact of dust deposition on solar photovoltaic systems: A comprehensive review of performance degradation, regional variations, and mitigation strategies. *Sustainability*, 18(10):4893, 2026.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 2017.

Wenpeng Miao. Bias correction of remote-sensed surface solar radiation and analysis of meteorological factor influences in plateau regions: A case study of lhasa. *Sustainability*, 18(12):6067, 2026.

B. O. Olorunfemi, Omolola Ogbolumani, and N. Nwulu. Solar panels dirt monitoring and cleaning for performance improvement: A systematic review on smart systems. *Sustainability*, 14(17): 10920, 2022.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, 2019.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 2021.

Hanane Rezzouk, Sarah Djabali, Ana-Maria Cretu, Shamsodin Taheri, Jamshidpour Ehsan, and Rezkallah Miloud. Characterization of photovoltaic modules under uniform snow covers. In *2025 IEEE Industry Applications Society Annual Meeting (IAS)*, pp. 1–5, 2025.

Wenyi Shi, Jingyi Yang, and Shaoyi Ren. Impact of dust accumulation on photovoltaic module power generation at a photovoltaic station in the zhongwei region. In *2025 4th International Conference on New Energy System and Power Engineering (NESP)*, pp. 1–4, 2025.

Jubayed Hossain Turag, Mohammad Amdadull Kabir, Md. Naim Al Hasan, and Ahmed Al Mansur. Soil impact on photovoltaic performance: A laboratory-based investigation. In *2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI)*, pp. 1–4, 2022.

Yihang Wu, Ahmad Chaddad, Christian Desrosiers, Tareef Daqqaq, and Reem Kateb. FAA-CLIP: Federated adversarial adaptation of CLIP. *IEEE Internet of Things Journal*, 12(12):21091–21102, 2025.

Zhimian Yin and Ciwei Gao. Federated learning-based anomaly classification detection for solar photovoltaic panels. In *2024 9th International Conference on Electronic Technology and Information Science (ICETIS)*, pp. 380–384, 2024.

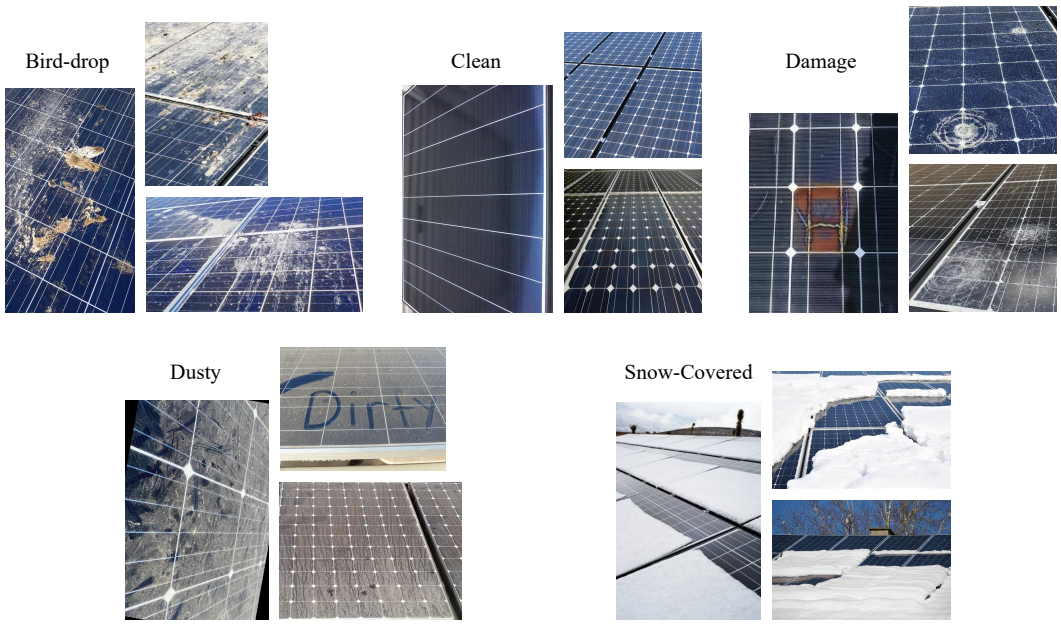

Figure 2: Samples from the dataset

A Dataset Details

Figure 2 presents representative sample images from the five soiling categories of the composite dataset.

Mild Heterogeneity (Simulating Realistic Conditions). This partitioning scheme simulates the real-world phenomenon whereby PV power plants at different geographic locations exhibit imbalanced distributions of soiling type samples due to differences in climatic conditions. It is worth noting that the clean state, as a non-soiling reference category, falls outside the scope of soiling distribution analysis; therefore, this study deliberately maintains a balanced distribution of 100 clean samples per site to ensure sufficient normal-state reference samples at each site. Under this configuration, although each site exhibits a certain degree of class skew, the overall sample distribution remains relatively balanced. For example, Site 2 (desert region) contains more dusty samples (300), while Site 3 (cold region) is dominated by snow-covered samples (95), consistent with real-world soiling distribution patterns across different climatic regions.

Severe Heterogeneity. This partitioning scheme further exacerbates inter-site class distribution disparities, simulating a more extreme data heterogeneity scenario. Under this configuration, each site is dominated by a single soiling type, and the resulting distributions exhibit artificially constructed characteristics that deviate from real-world conditions. For example, Site 1 (coastal region) is dominated by damage samples (370), Site 2 (desert region) by dusty samples (370), Site 3 (cold region) by snow-covered samples (117), and Site 4 (temperate plain) by bird-drop samples (370). Such extreme class skew places greater demands on the robustness of federated learning algorithms.

Tables 5 and 6 present the detailed sample counts for each site under the mild and severe heterogeneity partitioning schemes, respectively.

B Method Details

B.1 Domain-Specific Prompt Engineering

To address the domain-specific nature of PV panel soiling recognition, we introduce specialized prompt engineering at the text level. Departing from the generic template "`a photo of a {label}`", we construct text descriptions imbued with environmental priors for each power plant

Table 5: Partition scheme of the mildly heterogeneous dataset

Site	Bird-drop	Clean	Damage	Dusty	Snow-Covered	Total
Site 1 (Coastal)	220	100	120	30	10	480
Site 2 (Desert)	20	100	130	300	3	553
Site 3 (Cold)	10	100	60	20	95	325
Site 4 (Temperate)	150	100	90	50	15	405
Total	**400**	**400**	**400**	**400**	**123**	**1723**

Table 6: Partition scheme of the severely heterogeneous dataset

Site	Bird-drop	Clean	Damage	Dusty	Snow-Covered	Total
Site 1 (Coastal)	10	100	370	10	2	492
Site 2 (Desert)	10	100	10	370	2	492
Site 3 (Cold)	10	100	10	10	117	247
Site 4 (Temperate)	370	100	10	10	2	492
Total	**400**	**400**	**400**	**400**	**123**	**1723**

scenario. For instance, the prompt for a coastal site is formulated as `"a clean solar panel in a coastal humid environment"`, whereas that for a desert site is `"a dusty solar panel in a desert environment"`. These prompts explicitly inject domain expertise from PV O&M into the model initialization, endowing the text embeddings with scene awareness from the outset. In industrial practice, O&M personnel can flexibly adjust these natural language descriptions based on local experience. This enables rapid adaptation without the need to re-annotate data or modify the model architecture, fully leveraging the open-vocabulary recognition capability of CLIP.

B.2 Progressive Alpha Scheduling Strategy

To stabilize the scene adversarial training process, inspired by the sigmoid-based nonlinear scheduling in DANN, we design a simplified linear progressive scheduling strategy for the GRL scaling factor α. Specifically, α increases linearly from 0 to 1 during training:

$$\alpha(t) = \min\left(1.0, \frac{t}{0.5T}\right), \tag{9}$$

where t is the current iteration number and T is the total number of iterations. In the early stages of training, $\alpha \approx 0$, rendering the gradient reversal effect negligible. This allows the image adapter to first learn meaningful scene-specific features. As training proceeds, α gradually increases, strengthening the scene-confusion signal and progressively enforcing scene invariance. This linear scheduling strategy prevents the adversarial training from becoming unstable during the initial optimization phase.

B.3 Communication Protocol

Federated training proceeds in communication rounds. In each round: (1) each client performs local training for a specified number of iterations by minimizing Equation equation 8; (2) the client uploads its shared embedding $\mathbf{E}_k^s$ to the server; (3) the server aggregates the shared embeddings via the weighted average defined in Equation equation 6; and (4) the server broadcasts the aggregated global shared embedding to all clients, with no other model parameters transmitted.

This protocol highlights the communication efficiency of PFL. By transmitting only the lightweight text embeddings rather than the full model weights, the communication cost is substantially reduced. Meanwhile, the raw image data and scenario-specific parameters of each plant remain strictly local, thereby ensuring robust data privacy.

B.4 Industrial Deployment Perspective Analysis

The communication mechanism is particularly critical for PV power plants. Remote plants often rely on wireless cellular networks or satellite links, which offer limited and expensive bandwidth. Consequently, transmitting full deep network weights (typically on the order of hundreds of megabytes) is impractical in such environments. In contrast, the proposed method requires uploading only the lightweight shared text embeddings ($C \times d$, typically a few KB) per round. This drastically reduces the communication overhead, rendering federated collaborative training feasible even for plants located in deserts and plateaus with constrained infrastructure.

The privacy-preserving mechanism of pFedSolar helps dismantle data silos among PV power plant operators. Different subsidiaries within the same group, or competing plants in the same region, often refuse to share raw data due to commercial confidentiality and geographical sensitivities. The proposed framework enables operators to participate in collaborative modeling without leaking sensitive image data, providing a technical pathway for newly built plants to leverage the historical experience of mature plants during cold starts. For instance, a newly built cold plant, despite having scarce snow-covered samples, can acquire universal soiling knowledge from other plants through the federated sharing mechanism, thereby significantly reducing the data collection costs associated with initial model deployment.

C Implementation Details

All experiments utilize the CLIP ViT-B/32 backbone. We configure $K = 4$ clients, each corresponding to a distinct PV power plant scenario, with all clients participating in the training process. Prior to training, the image data are preprocessed using resizing to 256×256, center cropping to 224×224, and standard normalization. Additionally, we use the zero-shot inference of CLIP as a reference upper bound, which is obtained by directly applying the pre-trained CLIP model with hand-crafted prompt templates. Federated training is conducted for a total of 100 communication rounds, during which each client performs one round of local training per round. Optimization is performed using the Adam optimizer with a learning rate of 1×10^{-3}, momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.98$, and a weight decay of 0.2. The domain adversarial loss weight is set to $\lambda_{\text{dann}} = 0.1$, and the private embedding regularization coefficient is $\lambda_{\text{reg}} = 0.01$. For each experimental setting, we conduct five independent runs with random seeds 0, 42, 123, 456, and 789, and report both the average accuracy and the standard deviation. All experiments are implemented in PyTorch (Paszke et al., 2019) and executed on an NVIDIA RTX 4060 Ti GPU.

D Ablation Study Details

Tables 7 and 8 present the full per-seed ablation study results on Dataset 1 (mild heterogeneity) and Dataset 2 (severe heterogeneity), respectively.

Table 7: Ablation study results on Dataset 1 (mild heterogeneity). accuracy (%) is reported. v1: base; v2: +federated domain adversarial learning; v3: +decoupled text embedding learning; v4: +personalized federated aggregation and optimization strategy (complete method). **bold** indicates the best result.

	v1	v2	v3	v4
Seed 0	67.78	67.64	87.80	93.04
Seed 42	71.69	71.96	87.98	**93.04**
Seed 123	67.18	71.17	87.16	91.36
Seed 456	66.20	66.91	86.50	91.63
Seed 789	69.31	69.33	86.59	91.30
Avg	68.43±2.14	69.40±2.18	87.21±0.68	**92.07±0.89**

Table 8: Ablation study results on Dataset 2 (severe heterogeneity). accuracy (%) is reported. v1: base; v2: +federated domain adversarial learning; v3: +decoupled text embedding learning; v4: +personalized federated aggregation and optimization strategy (complete method). **bold** indicates the best result.

	v1	v2	v3	v4
Seed 0	58.42	60.46	77.81	92.86
Seed 42	66.84	65.31	84.18	**94.90**
Seed 123	57.65	59.19	80.61	92.60
Seed 456	60.46	60.97	82.40	94.39
Seed 789	58.42	57.66	79.34	93.37
Avg	60.36±3.77	60.72±2.87	80.87±2.50	**93.62±0.99**

E Additional Analysis

E.1 Generalization Capability Analysis

To further evaluate the localized performance of each algorithm across different clients (i.e., PV power plant scenarios), Figure 3 presents the accuracy comparison of four representative algorithms on the four clients under the random seed of 0.

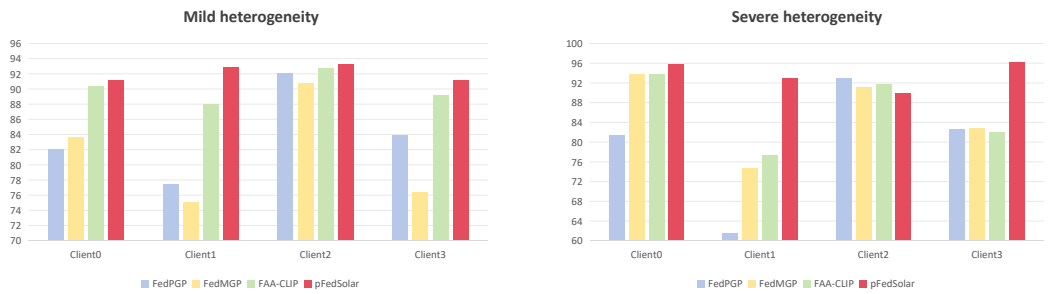

Figure 3: Performance comparison of each method across individual clients.

As illustrated in Figure 3, pFedSolar achieves optimal or near-optimal accuracy across all four clients, demonstrating outstanding generalization capability and local scene personalization. While FedPGP, FedMGP, and FAA-CLIP attain relatively high accuracy on certain clients, they still exhibit noticeable performance degradation on specific scenario-dependent clients. These results comprehensively validate that pFedSolar can effectively achieve personalized adaptation to local scenes while maintaining global knowledge generalization.

E.2 Hyperparameter Sensitivity Analysis

To evaluate the sensitivity of the proposed method to its core hyperparameters, we conduct a sensitivity analysis on the domain adversarial loss weight λ_{dann} and the private embedding regularization coefficient λ_{reg}. The experiments are performed separately on Dataset 1 (mild heterogeneity) and Dataset 2 (severe heterogeneity). During each evaluation, all other hyperparameters are kept fixed while only the target hyperparameter is varied. The results are presented in Table 9 and Table 10.

Table 9: Sensitivity analysis of the domain adversarial loss weight λ_{dann}. accuracy (%) is reported with $\lambda_{\text{reg}} = 0.01$. **bold** indicates the best result.

	0.01	0.05	0.1	0.2	0.5
Dataset 1 (Mild Heterogeneity)	92.69	92.73	**92.86**	92.77	92.56
Dataset 2 (Severe Heterogeneity)	92.89	92.92	**93.04**	92.89	92.78

Table 10: Sensitivity analysis of the private embedding regularization coefficient λ_{reg}. accuracy (%) is reported with $\lambda_{\text{dann}} = 0.1$. **bold** indicates the best result.

	0.001	0.005	0.01	0.02	0.05
Dataset 1 (Mild Heterogeneity)	92.71	92.79	**92.86**	92.74	92.68
Dataset 2 (Severe Heterogeneity)	92.88	92.92	**93.04**	92.86	92.79

As summarized in Tables 9 and 10, pFedSolar exhibits remarkably low sensitivity to both λ_{dann} and λ_{reg} across both datasets, with the accuracy fluctuation across all tested values not exceeding 0.3 percentage points. This robustness can be attributed to two factors: the structural design of the method and the distinct functional roles of each module.

The low sensitivity to λ_{dann} stems from the fact that the federated domain adversarial learning module serves an "auxiliary alignment" role at the representation level, rather than acting as the primary driver of performance improvement. The ablation study (Tables 7 and 8) reveals that transitioning from v1 to v2 yields only a modest gain of 0.36%–0.97%. This indicates that the visual features extracted by the CLIP pre-trained encoder already possess strong cross-domain invariance, and the distribution discrepancies among the four PV power plant scenarios (coastal, desert, high-latitude, and temperate) are inherently limited within the 512-dimensional CLIP embedding space. Furthermore, since the attention weighting mechanism of the image adapter A_k is already capable of capturing scene-specific visual differences, the marginal contribution of domain adversarial learning is naturally small. Consequently, varying λ_{dann} across a wide range (0.01 to 0.5) does not cause significant performance fluctuations. Notably, although the absolute contribution of DANN is modest, the substantial improvement from v2 to v3 (+15%–20%) in the ablation study highlights a synergistic effect between domain alignment and semantic decoupling: domain adversarial learning fosters scene invariance at the representation level, providing a more stable visual feature foundation for the subsequent decoupled text embedding learning.

The low sensitivity to λ_{reg} is primarily because the performance gains from private embeddings originate from a "hard constraint" at the aggregation strategy level (i.e., strictly keeping private embeddings local and excluding them from federated aggregation), rather than the "soft constraint" of regularization strength. The private embedding $\mathbf{E}_k^p$ is initialized as a zero vector and is driven by the classification loss to gradually learn scenario-specific semantic offsets, typically maintaining a small norm. Under these conditions, when λ_{reg} varies from 0.001 to 0.05, the absolute change in the regularization penalty is negligible relative to the gradient of the classification loss. The personalized adaptation capability of the private embedding fundamentally relies on the "no-aggregation" design choice, which ensures that each plant can independently preserve scenario-specific semantic knowledge without dilution from global averaging. In this context, λ_{reg} merely acts as a fine-tuning constraint to prevent semantic drift—analogous to weight decay—thus having a naturally limited impact on the final performance.

Furthermore, since the model already achieves an accuracy above 92% on both datasets, the optimization headroom is inherently limited, causing the marginal gains from hyperparameter variations to naturally plateau in this high-accuracy regime. In summary, the low sensitivity of pFedSolar to core hyperparameters reflects the strong robustness of its architectural design: each module achieves stable synergy through functional decoupling and differentiated aggregation strategies, maintaining consistent high performance without relying on fine-grained hyperparameter tuning. For PV power plant O&M teams, this low sensitivity implies that cross-plant batch deployment does not require professional deep learning engineers for per-site parameter tuning. Field technicians can simply adopt the default configuration to achieve stable, high performance across diverse environments. This "plug-and-play" feature significantly lowers the technical threshold for intelligent transformation and makes it possible for pFedSolar to be rapidly and widely deployed in the photovoltaic industry.

E.3 Confusion Matrix Analysis of Classification Results

To provide an in-depth analysis of the classification performance of pFedSolar across different soiling categories and identify potential misclassification patterns, Figures 4 and 5 present the normal-

ized confusion matrices of the proposed complete method on Dataset 1 (mild heterogeneity) and Dataset 2 (severe heterogeneity), respectively, with test samples aggregated across all clients.



Figure 4: Normalized confusion matrix on Dataset 1 (mild heterogeneity).

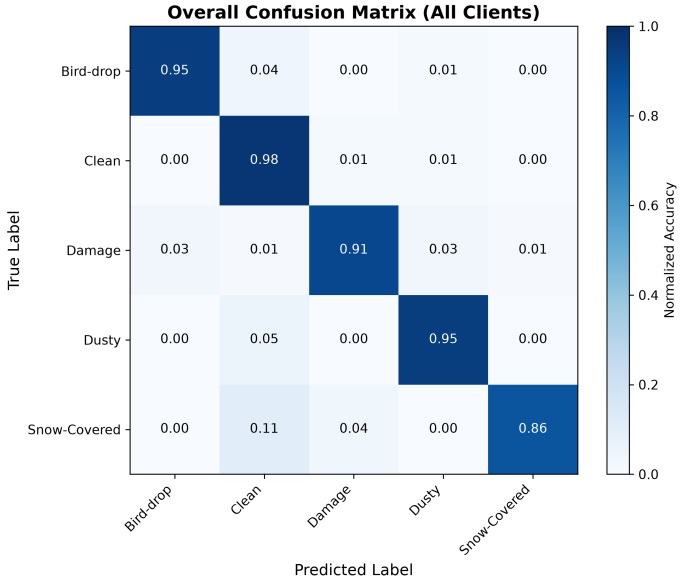

Figure 5: Normalized confusion matrix on Dataset 2 (severe heterogeneity).

As illustrated in Figure 4, pFedSolar achieves high classification accuracy across all five soiling categories under mild heterogeneity, as evidenced by the dominant diagonal elements. The snow-covered category attains a perfect recall of 1.00, with all samples correctly classified; this is attributed to the high-contrast textural features of snow, which are visually distinct from other soiling types. The recall rates for both bird-drop and clean categories are 0.92, demonstrating stable classification performance. Conversely, the dusty category exhibits the lowest recall (0.85), with misclassifications primarily directed toward physical/electrical damage (0.08) and clean (0.05). This occurs because the uniform deposition of dust on PV panel surfaces shares certain textural similarities with

mild physical damage, thereby posing a greater challenge for the model in distinguishing between these two categories.

As illustrated in Figure 5, under severe heterogeneity conditions, the classification accuracy for most categories actually improves: bird-drop increases from 0.92 to 0.95, clean from 0.92 to 0.98, dusty substantially increases from 0.85 to 0.95, and physical/electrical damage from 0.88 to 0.91. This trend aligns with the observations from our comparative experiments: greater data heterogeneity provides richer optimization signals for domain adversarial learning, enabling the model to learn more discriminative, scene-invariant features. The significant improvement in dusty accuracy (+0.10) is particularly noteworthy, indicating that the confusion between dusty and physical/electrical damage is effectively mitigated under severe heterogeneity, following a more thorough domain alignment of visual dust representations across clients. Conversely, the recall for snow-covered decreases from 1.00 to 0.86, with 11% of snow samples misclassified as clean. This degradation is primarily attributed to the limited number of snow samples in the severely heterogeneous dataset (only 123 images, exclusively from the high-latitude site), where the small sample size constrains the model's ability to adequately learn this category. Furthermore, since snow-covered and clean PV panels both exhibit high surface brightness, the feature boundary between them becomes increasingly difficult to precisely delineate when samples are scarce.

From an O&M economic perspective, specific misclassification patterns correspond directly to distinct financial risks. Misclassifying dusty panels as physical/electrical damage may trigger unnecessary component shutdowns and professional inspections, resulting in both labor costs and power generation revenue losses. Conversely, misclassifying snow-covered panels as clean may cause the winter O&M system to fail in promptly initiating snow removal operations, leading to continuous power generation losses during the cold season. By improving the dusty recall from 0.85 to 0.95 under severe heterogeneity, pFedSolar prevents approximately 10 misclassifications per 100 dusty cases. This directly translates into substantial O&M cost savings, demonstrating that the proposed method not only achieves SOTA-level technical metrics but also delivers quantifiable economic value for industrial deployment.