



FER-FL: Robust federated learning for facial expression recognition in heterogeneous environments

Yuanlun Xie^a, Jinshan Lai^b, Fengchun Zhang^b, Kaibo Shi^a, Deepak Kumar Jain^c, Vitomir Štruc^{d,*}, Nan Zhou^a, Badong Chen^e

^a School of Electronic Information and Electrical Engineering, Chengdu University, Chengdu, 610106, China

^b School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China

^c School of Computer Science, The University of Western Australia, Mumbai Campus, India

^d University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, 1000, Slovenia

^e Institute of Artificial Intelligence and Robotics (IAIR), Xi'an Jiaotong University, Shaanxi, Xian, 710049, China

ARTICLE INFO

Keywords:

Federated learning
Facial expression recognition
Privacy-aware learning

ABSTRACT

Vision-based facial expression recognition (FER) has shown strong potential across numerous applications, providing a non-invasive and real-time approach for analyzing emotional states and social intentions. However, deploying FER systems in practice often raises concerns about centralized facial data collection, as large-scale data from individual users or healthcare settings may contain sensitive personal information. Federated learning (FL), as a distributed training paradigm, reduces the need to share raw facial images by enabling collaborative model learning across decentralized clients. Despite this advantage, FL in the FER domain faces a major challenge: the non-independent and identically distributed (non-IID) nature of facial data, caused by variations in age, gender, ethnicity, identity, illumination, and individual expression styles. These discrepancies often lead to client drift, slow convergence, and reduced model accuracy. To address these issues, we propose FER-FL, a federated learning framework specifically designed for FER under heterogeneous client distributions. FER-FL introduces a client-side variance-reduced primal–dual optimization strategy to mitigate client drift and improve local–global consistency. In addition, a server-side public-data calibration mechanism is incorporated to refine the global decision boundary through KL-based distillation. Comprehensive experiments on FERPlus, RAF-DB, and FER2013 demonstrate that FER-FL improves Top-1 accuracy by up to 7.0% over FedAvg and accelerates convergence by up to 5.1×, while maintaining stable performance under different levels of data heterogeneity and client participation. The proposed framework provides an effective privacy-aware federated training solution for FER without requiring the exchange of raw facial images.

1. Introduction

Facial expression recognition (FER) aims to automatically infer human affective states from facial appearance and has emerged as a fundamental problem in computer vision and affective computing. Leveraging advances in deep learning, FER systems have demonstrated significant potential in applications such as intelligent monitoring, human–computer interaction, virtual assistants, healthcare support, and mental-state assessment [1–3]. Compared with contact-based sensing, vision-based FER provides a non-invasive, real-time solution for understanding emotional responses, social intentions, and behavioral cues, making it suitable for privacy-sensitive and resource-constrained scenarios. Despite these advances, deploying FER in real-world environments remains challenging due to the need to reduce centralized

collection of sensitive facial data, the inherent problem of data heterogeneity, and the robustness requirements of deep visual models in practical scenarios [4]. High-performance FER models typically require large-scale facial data containing personal information, which is difficult to centralize in applications such as healthcare or education due to regulatory and ethical considerations [5–8]. Furthermore, facial expressions are heavily influenced by age, gender, ethnicity, pose, illumination, and individual expression styles [9,10], resulting in datasets that are highly heterogeneous across different users and devices.

Federated learning (FL) provides a promising paradigm to reduce raw facial data sharing by enabling multiple clients to collaboratively train a shared model without exchanging their local images [11,12]. In this decentralized framework, only model parameters or updates are transmitted to a central server, while raw facial samples remain

* Corresponding author.

E-mail addresses: dkj@ieee.org (D.K. Jain), vitomir.struc@fe.uni-lj.si (V. Štruc).

<https://doi.org/10.1016/j.patcog.2026.114547>

Received 3 April 2026; Received in revised form 8 June 2026; Accepted 27 July 2026

Available online 24 August 2026

0031-3203/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

local, which aligns naturally with privacy-aware FER deployment [5, 13]. However, applying conventional FL algorithms directly to FER is often inadequate. Clients in federated FER scenarios typically present highly skewed label distributions and identity-specific facial patterns, resulting in significant local-global mismatch and the “client drift” problem [14,15]. Subtle inter-class differences, severe class imbalance, and subject-specific bias further exacerbate the training difficulty, causing standard aggregated global models to converge slowly, exhibit instability, or underperform on underrepresented emotional categories. Existing FL methods, including FedProx [16], SCAFFOLD [17], and FedDyn [18], do not explicitly account for these FER-specific challenges and lack effective global calibration mechanisms across heterogeneous expressions.

To address these gaps, we propose FER-FL, a federated learning framework specifically designed for facial expression recognition under heterogeneous client distributions. FER-FL integrates a client-side variance-reduced primal-dual update mechanism to reduce client drift and regulate local-global divergence, together with a server-side public-data calibration module to refine the global decision boundary. By combining these two components, FER-FL explicitly targets FER-specific challenges, including expression ambiguity, class imbalance, identity-sensitive bias, and subtle inter-class differences in raw facial data. The objective of this work is to improve federated FER under heterogeneous client data by jointly reducing client drift on the client side and calibrating the global decision boundary on the server side, without requiring the exchange of raw facial images. The key contributions of this work are summarized as follows:

- We formulate FER-FL as an FER-oriented federated optimization framework that addresses expression ambiguity, class imbalance, and identity-related distribution shifts under heterogeneous client data without sharing raw facial images;
- We introduce a variance-reduced primal-dual local update strategy, where corrected gradients and dual alignment jointly reduce client drift and improve local-global consistency during federated optimization;
- We design a public-data-calibrated global distillation mechanism to refine the aggregated decision boundary on the server side, and provide systematic evaluations including public-data-enhanced fairness baselines, public proxy data sensitivity analysis, convergence comparison, computational cost, and class-level performance.

The remainder of this paper is organized as follows. Section 2 reviews related work on FER and FL under heterogeneous data. Section 3 details the proposed FER-FL framework, including the client-side primal-dual optimization and server-side calibration mechanisms. Section 4 presents the experimental setup and a comprehensive analysis of the results, covering Top-1 accuracy, convergence speed, computational efficiency, granular performance via confusion matrices, and robustness under label noise. Finally, Section 6 concludes the paper and discusses potential future research directions.

2. Related work

This section reviews two lines of work closely related to FER-FL: federated learning under non-IID data and facial expression recognition. The former provides the distributed optimization background for handling heterogeneous clients, while the latter highlights the FER-specific challenges of expression ambiguity, class imbalance, identity bias, and raw-data constraints.

2.1. Federated learning with non-IID data

Federated learning (FL) enables multiple clients to collaboratively train a model without sharing raw data, and has been applied to

privacy-aware FER and healthcare-related scenarios [19]. However, client data are often non-IID due to differences in users, devices, environments, and data-collection protocols. Such heterogeneity can cause client drift and degrade the convergence and generalization of the global model [20–22].

A range of methods has been proposed to alleviate this problem. FedProx [16] constrains local updates with a proximal term, SCAFFOLD [17] uses control variates to correct client drift, and FedDyn [18] introduces dynamic regularization to improve local-global consistency. More recent methods, such as FedDC [23], FedSpeed [24], and FedSMO [25], further exploit gradient correction, variance reduction, or sharpness-aware optimization to improve convergence under heterogeneous data. FedNN also investigates federated learning under drifting distributions by reducing weight drift with normalization strategies [15].

Knowledge distillation has also been used to improve global consistency in FL. Representative methods include FedGKD [26], FedDF [27], and FedGen [14], which transfer knowledge through logits, ensembles, or generated representations. Nevertheless, most existing FL methods are designed for general classification tasks and do not explicitly address FER-specific issues such as expression ambiguity, long-tailed emotion categories, and identity-related bias. FER-FL addresses these limitations by combining variance-reduced primal-dual local optimization with server-side public-data calibration.

2.2. Facial expression recognition

Facial expression recognition (FER) aims to infer human emotions from facial cues and has been widely studied in computer vision and affective computing. Early studies established basic emotion categories and hand-crafted motion or appearance descriptors [28,29]. With the development of deep learning, FER has shifted toward data-driven representation learning, leading to stronger recognition performance in unconstrained environments [30].

Recent FER methods improve robustness through region-aware modeling, attention mechanisms, dynamic modeling, and specialized loss functions. Representative methods include RAN [31], M3DFEL [32], DDF [33], GCA+IAL [34], and PACVT [35]. Recent studies also focus on confusable FER in the wild, where low-quality images and subtle inter-class differences make recognition difficult [36]. Micro-expression recognition further emphasizes fine-grained and short-duration facial movements, and Micro_NesT improves such recognition with multi-scale attention [37].

Although these methods improve centralized FER performance, most of them assume that training data can be collected and processed centrally. In real applications, facial data are distributed across users or institutions and are affected by identity, age, gender, illumination, pose, and individual expression styles. These factors make federated FER more challenging than standard centralized FER. FER-FL therefore focuses on improving federated optimization for FER under heterogeneous client data while keeping raw facial images local.

3. The FER-FL framework

In this section, we present the proposed FER-FL framework for robust federated facial expression recognition under heterogeneous client distributions. The key idea is to jointly address two complementary sources of degradation in federated FER: local client drift caused by non-IID private data and global decision-boundary bias caused by class imbalance and expression ambiguity. To this end, FER-FL combines a variance-reduced primal-dual local optimization strategy on the client side with a public-data-based calibration module on the server side. The overall training process is summarized in Algorithm 1.

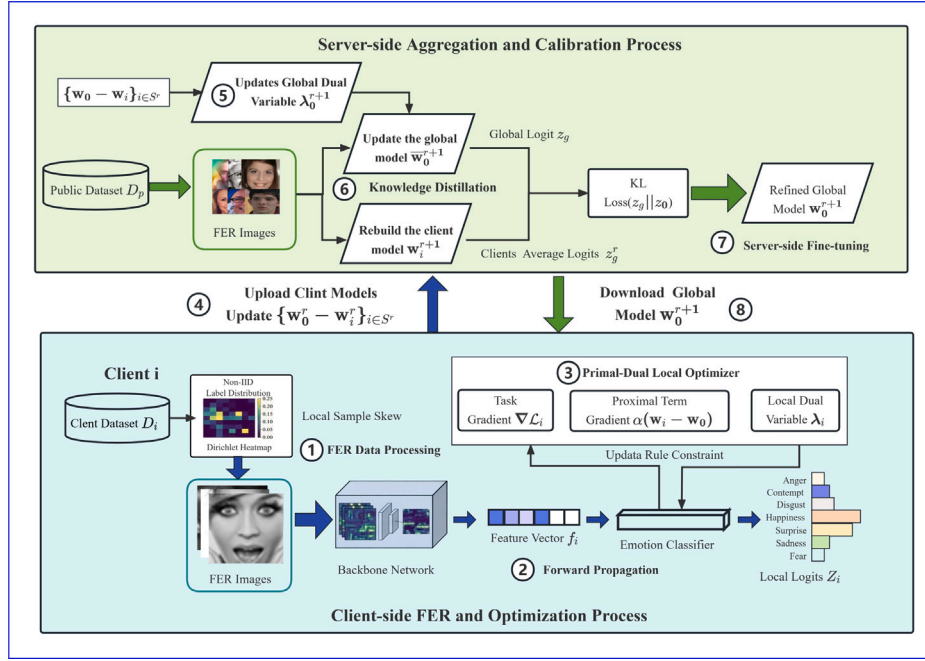


Fig. 1. Hierarchical architecture of the proposed FER-FL framework. The framework consists of distributed clients and a central server. Each communication round includes client-side local training on private datasets D_i using a primal-dual optimization strategy to extract emotional features while mitigating drift. The central server aggregates these local updates and performs global calibration on a public dataset D_p via knowledge distillation to refine the global model parameters.

3.1. Problem formulation

We consider a standard federated learning setting with one central server and N distributed clients. Each client i owns a private FER dataset D_i , which is not shared with the server or other clients. The server maintains a global model with parameters w_0 , while each client maintains a local model with parameters w_i . The local empirical risk of client i is defined as

$$f_i(w_i) = \frac{1}{|D_i|} \sum_{(x,y) \in D_i} \ell(w_i; x, y), \quad (1)$$

where $\ell(\cdot)$ denotes the FER classification loss, such as cross-entropy.

The objective of federated FER is to learn a global model that generalizes across heterogeneous clients without uploading raw facial images:

$$\min_{w_0} F(w_0) = \frac{1}{N} \sum_{i=1}^N f_i(w_0). \quad (2)$$

However, FER data are naturally heterogeneous due to identity-related facial patterns, expression styles, age, gender, illumination, pose, and class imbalance. As a result, local optimization on each client may deviate from the global objective, leading to client drift and unstable aggregation. FER-FL addresses this issue through client-side variance-reduced primal-dual optimization and server-side public-data calibration.

3.2. Overview of FER-FL

As illustrated in Fig. 1, FER-FL follows an iterative client-server training procedure. In each communication round r , the server selects a subset of clients S^r and broadcasts the current global model w_0^r to them. Each selected client initializes its local model with w_0^r and performs local training on its private FER data.

The local training stage contains two drift-control components. First, a gradient correction term is introduced to reduce the discrepancy between each client's historical update direction and the average

update direction. Second, a client-specific dual variable λ_i is maintained to penalize the inconsistency between the local model and the global model. These two components jointly encourage local updates to remain aligned with the global optimization objective.

After local training, each selected client sends its model update to the server. The server reconstructs the updated local models, aggregates them into a preliminary global model, and then performs public-data-based calibration using a small disjoint proxy dataset D_p . Specifically, the server computes the prediction consensus of local models on D_p and fine-tunes the aggregated global model by minimizing the KL divergence between the global prediction and the client consensus. The calibrated model is then used as the global model for the next communication round.

3.3. Variance-reduced primal-dual local optimization

At the beginning of round r , the server broadcasts w_0^r to each selected client $i \in S^r$. The client initializes

$$w_i^{r,0} = w_0^r. \quad (3)$$

To reduce update inconsistency under non-IID data, we introduce a variance-reduced gradient correction term. Let Δ_i^{r-1} denote the historical update of client i from the previous round in which it participated, and let $\bar{\Delta}^{r-1}$ denote the mean update over the participating clients in the previous communication round:

$$\bar{\Delta}^{r-1} = \frac{1}{|S^{r-1}|} \sum_{j \in S^{r-1}} \Delta_j^{r-1}. \quad (4)$$

We define the correction function as

$$G_i(w_i; \Delta_i^{r-1}, \bar{\Delta}^{r-1}) = \frac{1}{\eta T} \left\langle w_i, \Delta_i^{r-1} - \bar{\Delta}^{r-1} \right\rangle, \quad (5)$$

where η is the local learning rate and T is the number of local update steps. The corresponding gradient correction term is

$$g_{i,1}^r = \nabla_{w_i} G_i(w_i; \Delta_i^{r-1}, \bar{\Delta}^{r-1}) = \frac{1}{\eta T} (\Delta_i^{r-1} - \bar{\Delta}^{r-1}). \quad (6)$$

For clients without historical updates, $g_{i,1}^r$ is initialized to zero.

To further constrain local-global divergence, FER-FL introduces a client-specific dual variable λ_i^r . The local primal-dual objective is formulated as

$$F_i(\mathbf{w}_i) = f_i(\mathbf{w}_i) + \langle \lambda_i^r, \mathbf{w}_i - \mathbf{w}_0^r \rangle + \frac{\alpha}{2} \|\mathbf{w}_i - \mathbf{w}_0^r\|^2, \quad (7)$$

where α controls the strength of the proximal regularization. The dual term encourages the local model to remain consistent with the global reference model, while the proximal term explicitly penalizes excessive local deviation.

At local step t , client i samples a mini-batch $\xi_i^{r,t}$ from D_i and computes the stochastic gradient

$$\mathbf{g}_i^{r,t} = \nabla f_i(\mathbf{w}_i^{r,t}; \xi_i^{r,t}). \quad (8)$$

The corrected gradient is then defined as

$$\hat{\mathbf{g}}_i^{r,t} = \mathbf{g}_i^{r,t} + \mathbf{g}_{i,1}^r. \quad (9)$$

The primal update is given by

$$\mathbf{w}_i^{r,t+1} = \mathbf{w}_i^{r,t} - \eta [\hat{\mathbf{g}}_i^{r,t} + \alpha (\mathbf{w}_i^{r,t} - \mathbf{w}_0^r) + \lambda_i^r]. \quad (10)$$

This equation is consistent with Algorithm 1: the corrected gradient, the proximal regularizer, and the dual variable all participate in the local primal update.

After T local steps, the client obtains

$$\mathbf{w}_i^{r+1} = \mathbf{w}_i^{r,T}. \quad (11)$$

The client-side dual variable is updated as

$$\lambda_i^{r+1} = \lambda_i^r + \rho (\mathbf{w}_i^{r+1} - \mathbf{w}_0^r), \quad (12)$$

where ρ is the dual update step size. In our implementation, we set $\rho = 1$ unless otherwise specified. For non-participating clients, both their local model states and dual variables remain unchanged in the current round.

Finally, client i sends the model difference

$$\Delta_i^r = \mathbf{w}_i^{r+1} - \mathbf{w}_0^r \quad (13)$$

to the server. No raw facial images are transmitted.

3.4. Server aggregation and public-data calibration

After receiving the updates from selected clients, the server first reconstructs the updated local models:

$$\mathbf{w}_i^{r+1} = \mathbf{w}_0^r + \Delta_i^r, \quad i \in S^r. \quad (14)$$

The preliminary global model is obtained by averaging the updated local models:

$$\bar{\mathbf{w}}_0^{r+1} = \mathbf{w}_0^r + \frac{1}{|S^r|} \sum_{i \in S^r} \Delta_i^r = \frac{1}{|S^r|} \sum_{i \in S^r} \mathbf{w}_i^{r+1}. \quad (15)$$

Equivalently, this can be written as

$$\bar{\mathbf{w}}_0^{r+1} = \mathbf{w}_0^r + \frac{1}{|S^r|} \sum_{i \in S^r} (\mathbf{w}_i^{r+1} - \mathbf{w}_0^r). \quad (16)$$

This formulation fixes the sign inconsistency in the aggregation rule and ensures that the server moves the global model toward the average of the updated local models.

To calibrate the global decision boundary, the server then uses a small public proxy dataset D_p . The public dataset is used only on the server side and is disjoint from the private client training data. For each selected client model, the server computes the softened prediction on D_p :

$$\mathbf{p}_i^r = \text{softmax} \left(\frac{h(\mathbf{x}; \mathbf{w}_i^{r+1})}{\tau} \right), \quad \mathbf{x} \in D_p, \quad (17)$$

where $h(\cdot; \mathbf{w})$ denotes the model logits and τ is the distillation temperature. The server aggregates client predictions to form the ensemble consensus:

$$\mathbf{p}_g^r = \frac{1}{|S^r|} \sum_{i \in S^r} \mathbf{p}_i^r. \quad (18)$$

The prediction of the preliminary global model is computed as

$$\mathbf{p}_0^r = \text{softmax} \left(\frac{h(\mathbf{x}; \bar{\mathbf{w}}_0^{r+1})}{\tau} \right). \quad (19)$$

The server fine-tunes the preliminary global model by minimizing the KL divergence between the client consensus and the global prediction:

$$\mathcal{L}_{\text{KL}} = \text{KL}(\mathbf{p}_g^r \parallel \mathbf{p}_0^r) = \sum_{k=1}^K \mathbf{p}_g^r[k] \log \frac{\mathbf{p}_g^r[k]}{\mathbf{p}_0^r[k]}, \quad (20)$$

where K is the number of emotion classes. After T_1 fine-tuning epochs on D_p , the calibrated global model is denoted as

$$\mathbf{w}_0^{r+1} = \text{Calibrate}(\bar{\mathbf{w}}_0^{r+1}; D_p, \mathcal{L}_{\text{KL}}). \quad (21)$$

The calibrated model $\mathbf{w}_0^{r+1}$ is then broadcast to clients in the next communication round.

Algorithm 1: The FER-FL Framework

Input : Client datasets $\{D_i\}_{i=1}^N$, public proxy dataset D_p , rounds R , local steps T , server epochs T_1 , learning rate η , coefficients α, ρ , temperature τ , initial model $\mathbf{w}_0^0$, and $\lambda_i^0 = \mathbf{0}$.

Output: Final global model $\mathbf{w}_0^R$

```

1 Initialize  $\Delta_i^{-1} = \mathbf{0}$  for all clients;
2 for  $r = 0, 1, \dots, R-1$  do
3   Server selects  $S^r$  and broadcasts  $\mathbf{w}_0^r$ ;
4   Compute  $\bar{\Delta}^{r-1} = \frac{1}{|S^{r-1}|} \sum_{j \in S^{r-1}} \Delta_j^{r-1}$ ;
5   for each client  $i \in S^r$  in parallel do
6     Set  $\mathbf{w}_i^{r,0} = \mathbf{w}_0^r$  and  $\mathbf{g}_{i,1}^r = \frac{1}{\eta T} (\Delta_i^{r-1} - \bar{\Delta}^{r-1})$ ;
7     for  $t = 0, 1, \dots, T-1$  do
8       Compute  $\mathbf{g}_i^{r,t} = \nabla f_i(\mathbf{w}_i^{r,t}; \xi_i^{r,t})$ ;
9       Update  $\mathbf{w}_i^{r,t+1} = \mathbf{w}_i^{r,t} - \eta [\mathbf{g}_i^{r,t} + \mathbf{g}_{i,1}^r + \alpha (\mathbf{w}_i^{r,t} - \mathbf{w}_0^r) + \lambda_i^r]$ ;
10    end
11    Set  $\mathbf{w}_i^{r+1} = \mathbf{w}_i^{r,T}$ ;
12    Update  $\lambda_i^{r+1} = \lambda_i^r + \rho (\mathbf{w}_i^{r+1} - \mathbf{w}_0^r)$ ;
13    Send  $\Delta_i^r = \mathbf{w}_i^{r+1} - \mathbf{w}_0^r$  to the server;
14  end
15  Aggregate  $\bar{\mathbf{w}}_0^{r+1} = \mathbf{w}_0^r + \frac{1}{|S^r|} \sum_{i \in S^r} \Delta_i^r$ ;
16  Calibrate  $\bar{\mathbf{w}}_0^{r+1}$  on  $D_p$  for  $T_1$  epochs by minimizing  $\mathcal{L}_{\text{KL}}$  in Eq. (20);
17  Set the calibrated model as  $\mathbf{w}_0^{r+1}$ ;
18 end
```

4. Experimental setup

To evaluate the effectiveness and robustness of the proposed FER-FL framework, comprehensive experiments were conducted under various data distributions and participation settings. This section outlines the datasets and model architectures used, the baseline methods for comparison, and the implementation details.

4.1. Datasets

We evaluate FER-FL on three benchmark datasets: FERPlus [38], RAF-DB [39], and FER2013 [40]. We use their raw training and testing splits to retain real-world FER challenges, including class imbalance,

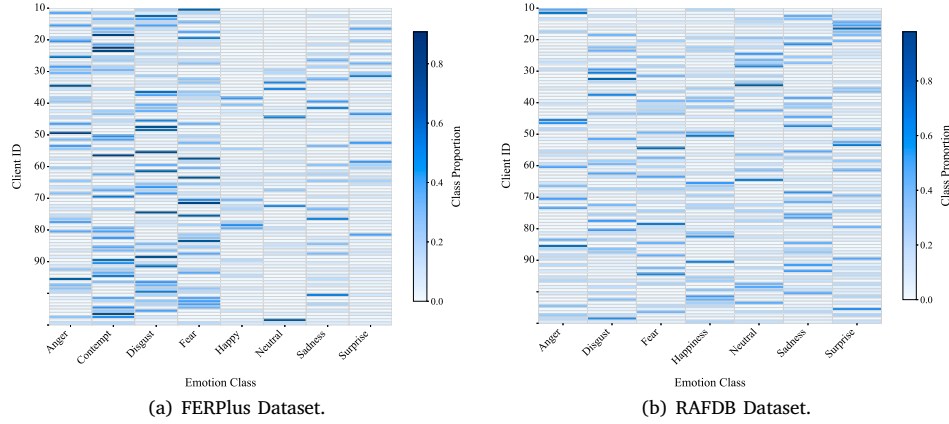


Fig. 2. Heat maps for the FERPlus and RAF-DB datasets under the D1 distribution. The figure illustrates the label distribution across 20 clients generated using the Dirichlet distribution with $\alpha = 0.3$, showing the degree of data heterogeneity among clients.

Table 1

Statistics of the public proxy dataset used for server-side calibration. The AffectNet subset is used only on the server and is disjoint from all client-side training and testing data.

Proxy dataset	#Samples	Classes	Label alignment
For FERPlus	4000	8 emotions, 500/class	Neutral, happiness, sadness, surprise, fear, disgust, anger, contempt
For RAF-DB	3500	7 emotions, 500/class	Neutral, happiness, sadness, surprise, fear, disgust, anger; contempt excluded
For FER2013	3500	7 emotions, 500/class	Neutral, happiness, sadness, surprise, fear, disgust, anger; contempt excluded

expression ambiguity, and long-tail label distributions. To simulate non-IID data, client-side samples are partitioned by a Dirichlet distribution $Dir(\alpha)$ with $\alpha = 0.3$ (D1) and $\alpha = 0.6$ (D2), where smaller α indicates stronger label heterogeneity. The D1 client distributions for FERPlus and RAF-DB are shown in Fig. 2.

For server-side calibration, FER-FL uses a small public proxy dataset D_p sampled from AffectNet [41], as summarized in Table 1. The proxy set is used only after server aggregation and is never involved in client-side training. It is also disjoint from the training and testing splits of FERPlus, RAF-DB, and FER2013. To ensure label-space consistency, we use eight AffectNet categories for FERPlus and remove the contempt category for RAF-DB and FER2013, which follow the seven-class setting. AffectNet labels are used only for balanced proxy sampling, while server calibration is performed by KL distillation from client logits rather than supervised cross-entropy on AffectNet labels.

4.2. Models

For the FERPlus and RAF-DB datasets, we adopt the ResNet-18 architecture as the expression recognition model, initialized with pre-trained ImageNet weights. We fine-tune the model on FERPlus and RAF-DB training data, respectively. For FER2013, we utilize a lightweight Convolutional Neural Network (CNN) consisting of two 5×5 convolutional layers and two 2×2 max-pooling layers, followed by two fully connected layers (120 and 84 units, respectively) with ReLU activation functions. These models are used to investigate the impact of heterogeneous data distributions on federated model performance, particularly in complex image classification tasks.

4.3. Baselines

We benchmark FER-FL against seven state-of-the-art FL methods: FedAvg [42], SCAFFOLD [17], FedDyn [18], FedCM [43], FedSpeed [24], FedGamma [44], and the most recent baseline FedGloss [45]. These baselines represent a cross-section of established and state-of-the-art federated-learning approaches and therefore allow for a comprehensive comparative evaluation of the proposed FER-FL framework.

4.4. Implementation details

FER-FL is implemented in PyTorch and trained on an NVIDIA GeForce RTX 4090 GPU. For fairness, all methods use batch size 50, $E = 5$ local updates, and 500 communication rounds. The local optimizer is SGD, with learning rate tuned from $\{0.01, 0.05, 0.1, 0.2\}$ and weight decay tuned from $\{1 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}\}$; the selected values are 0.1 and 1×10^{-3} . For FER-FL, α is tuned from $\{0.01, 0.05, 0.1, 0.2, 0.5\}$ and set to 0.1, while τ is tuned from $\{1, 2, 4, 6, 8\}$ and set to 4. Method-specific settings are listed in Table 2. FERPlus and RAF-DB are evaluated with $K = 20$ clients, while FER2013 uses $K = 50$ clients. Both 40% and 80% participation rates are tested. We report the mean and standard deviation over five seeds, i.e., 0, 1, 2, 3, 4. For each seed, the Dirichlet partition and client sampling order are regenerated. We use a label-wise Dirichlet protocol, where samples of each emotion class are distributed to clients according to $Dir(\alpha)$ with $\alpha = 0.3$ (D1) or $\alpha = 0.6$ (D2). The number of samples per client is kept balanced. The official training and testing splits are used, and evaluation is conducted on the official centralized test split with the global model. The AffectNet proxy subset is fixed across runs. FERPlus and RAF-DB images are face-cropped, resized to 224×224 , and normalized using ImageNet statistics. FER2013 images are resized to 48×48 and normalized as grayscale images. No additional face alignment is applied. Server-side calibration is performed on the AffectNet proxy subset for $T_1 = 5$ epochs using SGD, learning rate 0.01, batch size 50, and temperature $\tau = 4$. The server uses only KL distillation from selected client predictions, without any ground-truth cross-entropy loss on AffectNet labels.

5. Results and discussions

This section reports results that: (1) quantify the performance of the proposed FER-FL framework, (2) compare it against the selected state-of-the-art baseline methods, (3) evaluate its stability across varying non-IID conditions and client participation levels, and (4) examine its internal behavior via ablation and convergence analyses.

Table 2

Method-specific hyperparameter settings of FER-FL and baselines. Common training settings are described in the implementation details.

Method	Optimizer	Tuning range	Final selected value
FedAvg	SGD	–	–
SCAFFOLD	SGD	–	Control variates initialized as zero
FedDyn	SGD	{0.001, 0.01, 0.1}	Dynamic regularization coefficient = 0.01
FedCM	SGD	{0.01, 0.05, 0.1}	Correction coefficient = 0.05
FedSpeed	SAM-SGD	{0.01, 0.05, 0.1, 0.2}	SAM perturbation radius = 0.1
FedGamma	SAM-SGD	{0.01, 0.05, 0.1, 0.2}	SAM perturbation radius = 0.1
FedGloss	SAM-SGD	{0.01, 0.05, 0.1}	Global SAM perturbation radius = 0.01
FER-FL	SGD	$\alpha \in \{0.01, 0.05, 0.1, 0.2, 0.5\}; \tau \in \{1, 2, 4, 6, 8\}$	$\alpha = 0.1, \rho = 1, \tau = 4, T_1 = 5$

Table 3

Top-1 accuracy comparison on FERPlus, RAF-DB, and FER2013 datasets. The table presents the Top-1 classification accuracy (% $\pm$ SD) of FER-FL and baseline FL methods across different data distributions (D1, D2) and client participation rates (40% and 80%). Results are averaged over 5 independent runs, with the best results highlighted in bold.

Method	FERPlus		RAF-DB		FER2013
	D1	D2	D1	D2	IID
<i>40% participation</i>					
FedAvg	57.78 $\pm$ 0.42	59.98 $\pm$ 0.38	33.02 $\pm$ 0.55	41.09 $\pm$ 0.49	52.24 $\pm$ 0.31
SCAFFOLD	63.23 $\pm$ 0.25	65.54 $\pm$ 0.22	37.90 $\pm$ 0.31	45.67 $\pm$ 0.28	53.10 $\pm$ 0.24
FedDyn	59.12 $\pm$ 0.33	62.61 $\pm$ 0.31	35.65 $\pm$ 0.40	43.78 $\pm$ 0.35	52.84 $\pm$ 0.28
FedCM	62.76 $\pm$ 0.28	63.49 $\pm$ 0.26	36.97 $\pm$ 0.34	44.94 $\pm$ 0.30	51.50 $\pm$ 0.33
FedSpeed	63.57 $\pm$ 0.19	65.90 $\pm$ 0.15	37.89 $\pm$ 0.25	45.98 $\pm$ 0.22	56.11 $\pm$ 0.18
FedGamma	63.67 $\pm$ 0.18	66.43 $\pm$ 0.14	38.12 $\pm$ 0.21	46.09 $\pm$ 0.19	56.60 $\pm$ 0.15
FedGloss	63.63 $\pm$ 0.21	66.18 $\pm$ 0.17	38.05 $\pm$ 0.24	46.03 $\pm$ 0.20	56.38 $\pm$ 0.19
FER-FL (ours)	64.37 $\pm$ 0.12	66.79 $\pm$ 0.11	38.62 $\pm$ 0.18	46.41 $\pm$ 0.15	57.30 $\pm$ 0.12
<i>80% participation</i>					
FedAvg	60.34 $\pm$ 0.35	62.54 $\pm$ 0.31	34.65 $\pm$ 0.48	47.45 $\pm$ 0.42	54.10 $\pm$ 0.29
SCAFFOLD	64.13 $\pm$ 0.22	65.87 $\pm$ 0.19	38.41 $\pm$ 0.28	51.78 $\pm$ 0.25	55.10 $\pm$ 0.21
FedDyn	62.11 $\pm$ 0.30	63.87 $\pm$ 0.27	36.65 $\pm$ 0.37	50.09 $\pm$ 0.32	54.90 $\pm$ 0.25
FedCM	63.65 $\pm$ 0.26	64.13 $\pm$ 0.23	37.91 $\pm$ 0.31	50.93 $\pm$ 0.28	53.15 $\pm$ 0.28
FedSpeed	64.76 $\pm$ 0.16	65.89 $\pm$ 0.14	38.56 $\pm$ 0.22	51.87 $\pm$ 0.20	58.20 $\pm$ 0.17
FedGamma	64.98 $\pm$ 0.15	66.05 $\pm$ 0.13	39.04 $\pm$ 0.19	51.98 $\pm$ 0.17	58.70 $\pm$ 0.14
FedGloss	64.88 $\pm$ 0.18	66.01 $\pm$ 0.15	38.86 $\pm$ 0.21	51.93 $\pm$ 0.18	58.46 $\pm$ 0.16
FER-FL (ours)	65.12 $\pm$ 0.10	67.08 $\pm$ 0.09	39.57 $\pm$ 0.15	52.09 $\pm$ 0.14	59.00 $\pm$ 0.11

5.1. Comparison with state-of-the-art methods

The Top-1 accuracy results of FER-FL and seven representative FL baselines are reported in Table 3. Beyond absolute accuracy gains, we further examine whether FER-FL provides stable improvements under FER-specific challenges, including non-IID label distributions, raw facial images, class imbalance, and partial client participation.

5.1.1. Performance comparison

Under the 40% participation setting, only a subset of clients contributes to each communication round, making global optimization more vulnerable to client drift and biased local updates. In this challenging setting, FER-FL improves over FedAvg by 6.6%–7.0% on FER-Plus and 3.6%–5.3% on RAF-DB, indicating that the proposed primal-dual update and server-side calibration help stabilize federated FER training under limited client participation. When the participation rate increases to 80%, all methods benefit from richer client information and the performance gaps become smaller. FER-FL still achieves the best overall accuracy across datasets. Compared with strong baselines such as FedGamma and FedGloss, the improvement is moderate in some cases, around 0.3%–0.7%, but remains consistent across datasets, data distributions, and participation rates. This suggests that FER-FL provides stable gains rather than relying on a single favorable setting. The absolute accuracy on RAF-DB is lower than that on FERPlus, mainly because we use raw data to preserve realistic FER challenges. RAF-DB contains more complex in-the-wild variations, including pose, illumination, identity, and expression ambiguity, while the non-IID partition further amplifies local class imbalance. Therefore, the lower

RAF-DB accuracy reflects the difficulty of federated FER under realistic raw-data distributions. In this setting, FER-FL still yields consistent improvements, showing its practical value for heterogeneous FER deployment.

5.1.2. Robustness under high heterogeneity

Under the more heterogeneous D1 setting, client data are more strongly skewed by label distribution, making local models easier to overfit to client-specific expression patterns. FER-FL maintains the best or highly competitive performance with smaller standard deviations across repeated runs, indicating improved training stability. This stability is important for FER applications because expression categories are often imbalanced and semantically close, where unstable aggregation may easily bias the global model toward majority or easier classes. The results in Table 3 therefore show that FER-FL not only improves Top-1 accuracy, but also provides a more reliable optimization process under non-IID FER data.

5.2. Evaluation of convergence and computational efficiency

Table 4 reports the number of communication rounds required to reach the target accuracy, while Table 5 further reports the wall-clock cost on FERPlus D2 with 40% participation. These results complement the Top-1 accuracy comparison in Table 3 by showing whether the accuracy gains are obtained efficiently.

Communication rounds. FER-FL consistently requires fewer communication rounds than the baselines to reach the target accuracy. For example, on FERPlus D2 with 40% participation, FER-FL reaches the

Table 4

Convergence comparison under D2 (FERPlus, RAF-DB) and IID (FER2013) settings. The table reports the number of communication rounds ($R\#$) required by each method to reach the target accuracy and the relative convergence rate ($S \uparrow$) with respect to FedAvg. The symbol > denotes methods that fail to reach the target accuracy within 500 rounds.

Method	FERPlus (Target = 60%, D2)				RAF-DB (Target = 35%, D2)				FER2013 (Target = 55%, IID)			
	40% part.		80% part.		40% part.		80% part.		40% part.		80% part.	
	$R\#$	$S \uparrow$	$R\#$	$S \uparrow$	$R\#$	$S \uparrow$	$R\#$	$S \uparrow$	$R\#$	$S \uparrow$	$R\#$	$S \uparrow$
FedAvg	>500	1.00	467	1.00	478	1.00	488	1.00	425	1.00	380	1.00
SCAFFOLD	276	1.81	266	1.76	134	3.57	117	4.17	210	2.02	185	2.05
FedDyn	467	1.07	442	1.06	176	2.72	169	2.89	315	1.35	295	1.29
FedCM	312	1.60	231	2.02	187	2.56	169	2.89	255	1.67	210	1.81
FedSpeed	234	2.14	232	2.01	123	3.89	119	4.10	165	2.58	142	2.68
FedGamma	217	2.30	185	2.52	124	3.85	109	4.48	148	2.87	128	2.97
FedGloss	228	2.19	225	2.08	118	4.05	115	4.24	155	2.74	135	2.81
FER-FL (ours)	165	3.03	124	3.77	97	4.93	87	5.61	105	4.05	82	4.63

Table 5

Comparison of computational overhead and training efficiency. We report the average training time per communication round (s) and the total time (h) required to reach the target accuracy on the FERPlus dataset (Target = 60%, D2 setting, 40% participation).

Method	Time per round (s)	Rounds to target (R_g)	Total time (h)	Speedup ($S \uparrow$)
FedAvg	0.72 ± 0.12	>500	>0.10	1.00×
SCAFFOLD	0.75 ± 0.13	276	0.057	1.75×
FedDyn	0.88 ± 0.13	467	0.114	0.88×
FedSpeed	1.29 ± 0.24	234	0.084	1.19×
FedGamma	1.38 ± 0.23	217	0.083	1.20×
FedGloss	1.75 ± 0.24	228	0.111	0.90×
FER-FL (Ours)	1.24 ± 0.15	165	0.056	1.78×

Table 6

Per-class accuracy (%) comparison on the FERPlus dataset (8 classes, D2 setting). This table demonstrates FER-FL's performance across the expanded emotion model. Despite the balanced sample distribution, the framework significantly improves the recognition of semantically complex categories like 'Contempt' and 'Fear'.

Method	Anger	Cont.	Disg.	Fear	Happy	Neut.	Sad.	Surp.	Macro-F1
FedAvg	59.8	52.4	54.1	50.3	76.2	61.5	55.4	62.1	59.0
SCAFFOLD	64.2	56.8	58.5	54.2	80.5	66.4	60.2	66.5	63.4
FedDyn	61.5	53.2	56.4	52.8	78.4	64.2	58.1	63.8	61.1
FedCM	62.8	55.1	57.2	53.5	79.1	65.3	59.4	64.7	62.2
FedSpeed	65.4	58.2	60.1	55.8	81.6	67.5	61.4	68.1	64.8
FedGamma	66.2	58.9	60.8	56.5	82.4	68.2	62.1	68.9	65.5
FedGloss	65.8	58.5	60.4	56.1	81.9	67.8	61.8	68.5	65.1
FER-FL (Ours)	67.2	61.5	62.8	59.4	84.5	70.4	64.5	71.2	67.7

60% target in 165 rounds, while FedAvg fails to reach the target within 500 rounds and SCAFFOLD requires 276 rounds. Similar trends are observed on RAF-DB and FER2013. This faster convergence is meaningful for federated FER because communication is often constrained in privacy-sensitive or edge-device scenarios. The result suggests that reducing client drift through primal-dual regularization and gradient correction can improve not only final accuracy but also training efficiency. On RAF-DB, FER-FL also reaches the target accuracy with the fewest rounds, although the absolute target is lower than that of FERPlus. This again reflects the higher difficulty of raw RAF-DB under non-IID partitioning. The convergence advantage shows that FER-FL can still provide stable optimization even when the underlying FER task is more complex and the data distribution is less favorable.

Wall-clock efficiency. Table 5 shows that FER-FL introduces moderate per-round overhead compared with FedAvg and SCAFFOLD, mainly due to the primal-dual update and server-side calibration. However, this overhead is lower than or comparable to SAM-based baselines such as FedSpeed, FedGamma, and FedGloss. More importantly, because FER-FL substantially reduces the number of required communication rounds, it achieves the shortest total training time to reach the target accuracy. This indicates that the proposed design provides a favorable trade-off between optimization stability, communication efficiency, and computational cost.

5.3. Granular performance analysis on raw data distributions

To further examine class-level performance, we report the per-class accuracy on FERPlus in Table 6 and visualize the confusion patterns on three datasets in Fig. 3. Since raw FER datasets often contain class imbalance and semantically similar expressions, global Top-1 accuracy alone may not fully reflect the model's behavior on difficult emotion categories. FER-FL improves the recognition of challenging categories such as Contempt and Fear, achieving accuracies of 61.5% and 59.4%, respectively. This indicates that server-side calibration helps reduce the bias of the global model toward dominant classes such as Neutral and Happiness. In addition, the confusion matrices show that FER-FL better distinguishes semantically similar expressions, such as Fear and Surprise, suggesting that the variance-reduced primal-dual update improves local feature stability under long-tailed and heterogeneous data. FER-FL also achieves the highest Macro-F1 score of 67.7% on FERPlus, showing that its improvement is not limited to majority classes. Instead, the proposed framework provides more balanced recognition across the emotion categories, which is important for practical FER applications where minority emotional states should not be ignored.

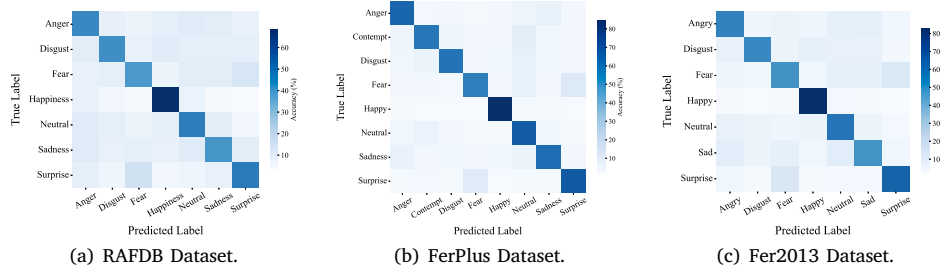


Fig. 3. Confusion matrices of FER-FL on three benchmark datasets. The diagonal elements represent the recall for each emotion category.

Table 7

Fairness comparison with public-data-enhanced baselines. All KD-enhanced baselines use the same AffectNet proxy subset, distillation temperature, and server fine-tuning epochs as FER-FL. Results are reported under FERPlus D2, RAF-DB D2, and FER2013 IID with 40% client participation.

Method	FERPlus D2	RAF-DB D2	FER2013 IID
FedAvg	59.98 $\pm$ 0.38	41.09 $\pm$ 0.49	52.24 $\pm$ 0.31
FedAvg+KD	61.25 $\pm$ 0.31	42.89 $\pm$ 0.38	53.87 $\pm$ 0.34
SCAFFOLD	65.54 $\pm$ 0.22	45.67 $\pm$ 0.28	53.10 $\pm$ 0.24
SCAFFOLD+KD	65.93 $\pm$ 0.25	46.11 $\pm$ 0.23	54.25 $\pm$ 0.21
FedDyn	62.61 $\pm$ 0.31	43.78 $\pm$ 0.35	52.84 $\pm$ 0.28
FedDyn+KD	63.58 $\pm$ 0.35	44.28 $\pm$ 0.28	53.47 $\pm$ 0.34
FER-FL (ours)	66.79 $\pm$ 0.11	46.41 $\pm$ 0.15	57.30 $\pm$ 0.12

Table 8

Sensitivity analysis on the size of the public proxy dataset. We vary the proportion of the AffectNet proxy subset used for server-side calibration and report Top-1 accuracy (%) and the number of communication rounds required to reach the target accuracy. Results are reported under FERPlus D2 with 40% client participation.

Proxy data size	Top-1 accuracy (%)	Rounds to target
5%	63.45 $\pm$ 0.21	285
10%	65.25 $\pm$ 0.23	267
25%	65.58 $\pm$ 0.18	243
50%	66.44 $\pm$ 0.15	196
100%	66.79 $\pm$ 0.11	165

5.4. Analysis of public proxy data

Since FER-FL uses an AffectNet proxy subset for server-side calibration, we conduct two additional analyses to examine whether the improvement comes solely from public data. First, we compare FER-FL with public-data-enhanced baselines, including FedAvg+KD, SCAFFOLD+KD, and FedDyn+KD. These baselines use the same proxy subset, distillation temperature, and server fine-tuning epochs as FER-FL. As shown in Table 7, KD calibration improves all three baselines, confirming that the proxy data provide useful calibration signals. However, FER-FL still achieves the best performance, improving over the strongest KD-enhanced baseline by 0.86%, 0.30%, and 3.05% on FER-Plus D2, RAF-DB D2, and FER2013 IID, respectively. This shows that the gain of FER-FL is not solely due to public data, but also comes from the proposed primal-dual optimization and gradient correction.

Second, we study the sensitivity to proxy data size in Table 8. As the proxy subset increases from 5% to 100%, accuracy improves from 63.45% to 66.79%, and the rounds to target decrease from 285 to 165. The improvement from 50% to 100% is relatively small, suggesting diminishing returns. These results indicate that FER-FL benefits from public calibration but does not heavily depend on a large-scale public dataset.

5.5. Sensitivity analysis

5.5.1. Impact of participation rates and local intervals

We first analyze the sensitivity of FER-FL to client participation rates and local training intervals. For the participation-rate experiments, the local interval is fixed at 5, while the participation rate varies in {0.2, 0.4, 0.6, 0.8, 1.0}, as shown in Fig. 4. A higher participation rate indicates that more clients contribute to each communication round. As expected, increasing the participation rate generally improves accuracy because the server receives more diverse client updates. However, the improvement becomes marginal when the participation rate increases from 0.8 to 1.0, suggesting that FER-FL can obtain sufficient global information under high but partial participation.

For the local-interval sensitivity experiment, we additionally consider a larger setting with 100 clients and fix the participation rate to 40%. The local intervals are selected from {1, 2, 5, 10, 20}, corresponding to 2500, 1250, 500, 250, and 125 global communication rounds, respectively, as shown in Fig. 5. Larger local intervals reduce communication frequency but may also increase local overfitting and client drift. Therefore, an appropriate local interval should balance communication efficiency and global generalization. FER-FL maintains relatively stable performance under different local intervals, showing that the proposed drift-control mechanisms remain effective even when local training becomes longer.

5.5.2. Hyperparameter sensitivity

Table 9 shows the influence of the proximal coefficient α and the distillation temperature τ on FER-FL across FERPlus, RAF-DB, and FER2013. The proximal coefficient α is selected from {0.01, 0.05, 0.1, 0.2, 0.5}, while τ is selected from {1, 2, 4, 6, 8}.

The results show that moderate values of both parameters lead to the best performance. Specifically, $\alpha = 0.1$ achieves the highest accuracy on all three datasets, indicating that a moderate proximal constraint is sufficient to reduce local-global divergence without overly restricting local adaptation. Similarly, $\tau = 4$ consistently produces the best results, suggesting that an intermediate distillation temperature provides informative soft targets for server-side calibration. Very small or very large values of α and τ lead to lower accuracy, either because the regularization/calibration effect is too weak or because the model becomes over-constrained.

5.6. Ablation study

To align the ablation settings with the proposed formulation and Algorithm 1, we evaluate three variants. “w/o GC” removes the gradient correction term in Eqs. (6) and (10). “w/o DR” removes the dual regularization by setting $\lambda_i^r = 0$ in Eq. (10) and disabling Eq. (12). “w/o Public Data” removes the server-side public calibration in Eqs. (17)–(20), so the server only performs aggregation in Eq. (15). As shown in Table 10, removing any component consistently degrades performance. On FERPlus D2, removing public calibration reduces accuracy from 66.79% to 63.05%, showing the benefit of AffectNet-based global cal-

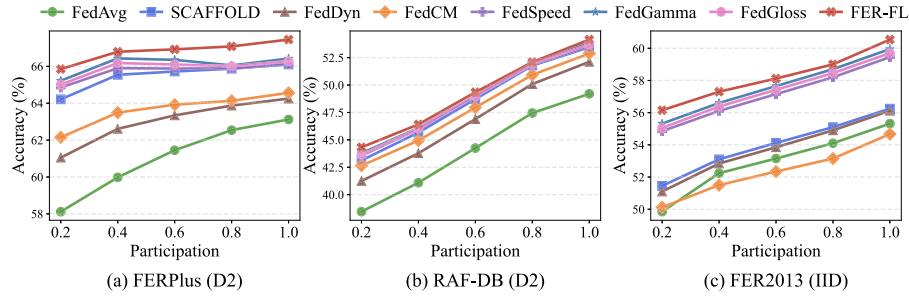


Fig. 4. Top-1 accuracy of FER-FL under different participation rates on RAF-DB, FERPlus, and FER2013 datasets.

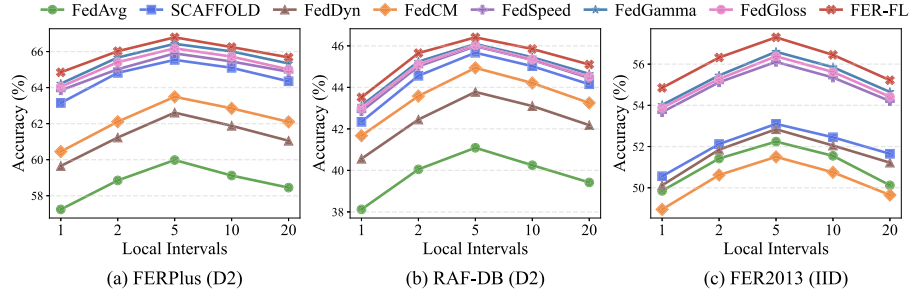


Fig. 5. Top-1 accuracy of FER-FL with different local intervals on RAF-DB, FERPlus, and FER2013 datasets.

Table 9

Impact of the proximal term α and distillation temperature τ in FER-FL. The table shows the effect of varying α and τ on Top-1 accuracy (% $\pm$ SD) across three datasets. Optimal values are highlighted in bold, representing the mean results over 5 independent runs.

α	0.01	0.05	0.10	0.20	0.50
FERPlus	60.65 $\pm$ 0.35	62.12 $\pm$ 0.28	63.04 $\pm$ 0.15	62.76 $\pm$ 0.24	60.12 $\pm$ 0.42
RAF-DB	34.13 $\pm$ 0.48	36.98 $\pm$ 0.39	38.17 $\pm$ 0.22	37.32 $\pm$ 0.31	35.34 $\pm$ 0.55
FER2013	56.02 $\pm$ 0.25	56.78 $\pm$ 0.20	57.30 $\pm$ 0.12	57.09 $\pm$ 0.18	56.12 $\pm$ 0.33
τ	1	2	4	6	8
FERPlus	61.43 $\pm$ 0.32	62.65 $\pm$ 0.25	63.04 $\pm$ 0.15	62.76 $\pm$ 0.21	62.09 $\pm$ 0.30
RAF-DB	36.21 $\pm$ 0.45	37.43 $\pm$ 0.36	38.17 $\pm$ 0.22	37.65 $\pm$ 0.28	37.24 $\pm$ 0.40
FER2013	56.20 $\pm$ 0.22	56.88 $\pm$ 0.18	57.30 $\pm$ 0.12	56.23 $\pm$ 0.26	55.34 $\pm$ 0.38

Table 10

Ablation study on the FER-FL framework across three datasets. The table reports Top-1 accuracy (% $\pm$ SD) of the full FER-FL model and its variants: without gradient correction (w/o GC), without dynamic regularization (w/o DR), and without server-side public calibration (w/o Public Data). Results are averaged over 5 independent runs.

Method variant	FERPlus		RAF-DB		FER2013
	D1	D2	D1	D2	IID
FER-FL w/o GC	62.45 $\pm$ 0.25	64.13 $\pm$ 0.21	36.76 $\pm$ 0.35	44.65 $\pm$ 0.29	55.48 $\pm$ 0.18
FER-FL w/o DR	57.40 $\pm$ 0.41	59.18 $\pm$ 0.35	30.98 $\pm$ 0.52	39.38 $\pm$ 0.44	51.25 $\pm$ 0.28
FER-FL w/o public data	61.22 $\pm$ 0.29	63.05 $\pm$ 0.24	35.10 $\pm$ 0.38	43.22 $\pm$ 0.31	54.85 $\pm$ 0.20
FER-FL (Full)	64.37 $\pm$ 0.12	66.79 $\pm$ 0.11	38.62 $\pm$ 0.18	46.41 $\pm$ 0.15	57.30 $\pm$ 0.12

Table 11

Ablation study with convergence analysis under FERPlus D2 and 40% client participation.

Method variant	Top-1 accuracy (%)	Rounds to target
FER-FL w/o GC	64.13 $\pm$ 0.21	238
FER-FL w/o DR	59.18 $\pm$ 0.35	285
FER-FL w/o Public Data	63.05 $\pm$ 0.24	259
FER-FL (Full)	66.79 $\pm$ 0.11	165

ibration. Removing GC decreases accuracy to 64.13%, indicating that gradient correction helps reduce client drift. The largest drop occurs when DR is removed, where accuracy falls to 59.18%, confirming the importance of primal–dual regularization for local–global consistency. Table 11 further shows that full FER-FL reaches the target accuracy in 165 rounds, faster than w/o GC, w/o DR, and w/o Public Data, which require 238, 285, and 259 rounds, respectively. These results demonstrate that the three components jointly improve both accuracy and convergence efficiency.

6. Conclusion

This paper presented FER-FL, a federated learning framework for facial expression recognition under heterogeneous client data. FER-FL combines variance-reduced primal-dual local optimization with server-side public-data calibration. The client-side update reduces drift caused by non-IID local facial data, while the server-side calibration refines the global decision boundary without requiring clients to share raw facial images. Extensive experiments on FERPlus, RAF-DB, and FER2013 show that FER-FL consistently improves Top-1 accuracy, convergence speed, computational efficiency, and class-level recognition performance. The fairness comparison with public-data-enhanced baselines and the public-data sensitivity analysis further show that the performance gain is not solely due to the use of public proxy data. For the FER and Pattern Recognition communities, FER-FL highlights the importance of addressing expression ambiguity, class imbalance, identity bias, and local-global misalignment in federated FER. At the same time, the study has several limitations. First, the server-side calibration relies on a public proxy dataset, and its effectiveness may decrease when the proxy data have severe domain mismatch or insufficient label coverage. Second, FER-FL does not by itself prevent information leakage from model updates. As in standard FL, gradients or parameter updates may still contain sensitive information, so real deployments should combine FER-FL with secure aggregation, differential privacy, update clipping, or encryption-based protocols. In addition, the client-side dual variables introduce extra local storage, and server-side calibration brings moderate additional computation. Although FER-FL reduces the total time to reach the target accuracy, further optimization is needed for resource-constrained devices. Finally, our experiments mainly focus on image-based FER. Future work will extend FER-FL to video-based FER, multimodal affect recognition, and personalized emotion modeling, while exploring lightweight pruning, sparse communication, and stronger privacy mechanisms.

CRedit authorship contribution statement

Yuanlun Xie: Software, Methodology, Formal analysis, Conceptualization. **Jinshan Lai:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **Fengchun Zhang:** Validation, Methodology, Formal analysis, Conceptualization. **Kaibo Shi:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition. **Deepak Kumar Jain:** Writing – review & editing, Supervision, Resources, Methodology. **Vitomir Štruc:** Writing – review & editing, Validation, Supervision, Conceptualization. **Nan Zhou:** Validation, Supervision, Resources, Formal analysis, Data curation. **Badong Chen:** Writing – review & editing, Supervision, Investigation, Conceptualization.

Declarations & ethics

The authors declare that:

- The submitted manuscript in whole or in parts is not under review in any other venue and has not been published in any form before
- We used Generative AI only to proof read the paper and improve the language

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used ChatGPT in order to proofread the manuscript, improve the language and presentation of the tables. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of Sichuan Province (No. 2025ZNSFSC1498) and 2023 Chengdu University of Information Technology Science and Technology Innovation Capability Enhancement Plan Innovation Team Key Project (No. KYTD202322) and ARIS Research Programme Metrology and Biometric Systems, under grant P2-0250.

Data availability

We used public datasets.

References

- [1] J. Mao, R. Xu, X. Yin, Y. Chang, B. Nie, A. Huang, Y. Wang, Poster++: A simpler and stronger facial expression recognition network, *Pattern Recognit.* (2024) 110951.
- [2] Z. Zhao, Q. Liu, F. Zhou, Robust lightweight facial expression recognition network with label distribution training, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, 2021, pp. 3510–3519.
- [3] C. Zheng, M. Mendieta, C. Chen, Poster: A pyramid cross-fusion transformer network for facial expression recognition, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3146–3155.
- [4] R. Liu, K. Lam, W. Zhou, S. Wu, J. Zhao, D. Hu, M. Gong, STBA: towards evaluating the robustness of DNNs for query-limited black-box scenario, *IEEE Trans. Multim.* 27 (2025) 2666–2681, [Online]. Available: <https://doi.org/10.1109/TMM.2025.3535328>.
- [5] A. Salman, C. Busso, Privacy preserving personalization for video facial expression recognition using federated learning, in: *Proceedings of the 2022 International Conference on Multimodal Interaction*, 2022, pp. 495–503.
- [6] X. Gong, L. Song, R. Vedula, A. Sharma, M. Zheng, B. Planche, A. Innanje, T. Chen, J. Yuan, D. Doermann, Z. Wu, Federated learning with privacy-preserving ensemble attention distillation, *IEEE Trans. Med. Imaging* 42 (7) (2023) 2057–2067.
- [7] Y. Zhang, Z. Li, X. Han, S. Ding, J. Li, J. Wang, S. Ying, J. Shi, Pseudo-data based self-supervised federated learning for classification of histopathological images, *IEEE Trans. Med. Imaging* 43 (3) (2024) 902–915.
- [8] R. Liu, Z. Lyu, W. Zhou, S. Ng, *D-Judge*: How far are we? Assessing the discrepancies between AI-synthesized and natural images through multimodal guidance, in: C. Gurrin, K. Schoeffmann, M. Zhang, L. Rossetto, S. Rudinac, D. Dang-Nguyen, W. Cheng, P. Chen, J. Benoit-Pineau (Eds.), *Proceedings of the 33rd ACM International Conference on Multimedia*, MM 2025, Dublin, Ireland, October 27–31, 2025, ACM, 2025, pp. 10797–10806.
- [9] A. Ullah, J. Wang, M.S. Anwar, A. Ahmad, S. Nazir, H.U. Khan, Z. Fei, Fusion of machine learning and privacy preserving for secure facial expression recognition, *Secur. Commun. Networks* 2021 (1) (2021) 6673992.
- [10] F. Xu, D. Ahmed-Aristizabal, L. Petersson, D. Wang, X. Li, Facial expression recognition with controlled privacy preservation and feature compensation, in: *2025 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, IEEE*, 2025, pp. 2611–2621.
- [11] S. Zhou, G.Y. Li, Federated learning via inexact ADMM, *IEEE Trans. Pattern Anal. Mach. Intell.* 45 (8) (2023) 9699–9708.
- [12] Y. Zhang, Y. Xu, S. Wei, Y. Wang, Y. Li, X. Shang, Doubly contrastive representation learning for federated image recognition, *Pattern Recognit.* 139 (2023) 109507.
- [13] D.G. Nair, J.J. Nair, K.J. Reddy, C.A. Narayana, A privacy preserving diagnostic collaboration framework for facial paralysis using federated learning, *Eng. Appl. Artif. Intell.* 116 (2022) 105476.
- [14] Data-free knowledge distillation via generator-free data generation for non-IID federated learning, *Neural Netw.* (ISSN: 0893-6080) 179 (2024) 106627.
- [15] M. Kang, S. Kim, K.H. Jin, E. Adeli, K.M. Pohl, S.H. Park, FedNN: Federated learning on concept drift data using weight and adaptive group normalizations, *Pattern Recognit.* 149 (2024) 110230.
- [16] L. Tian, A. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, V. Smith, Federated optimization in heterogeneous networks, 2018, *ArXiv: Learning*, ArXiv: Learning.
- [17] S. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, A. Suresh, SCAFFOLD: Stochastic controlled averaging for federated learning, in: *International Conference on Machine Learning*, International Conference on Machine Learning, 2020.

- [18] D. Acar, Y. Zhao, R. Matas, M. Mattina, P. Whatmough, V. Saligrama, Federated learning based on dynamic regularization, in: International Conference on Learning Representations, International Conference on Learning Representations, 2021.
- [19] L. Sun, J. Wu, A scalable and transferable federated learning system for classifying healthcare sensor data, *IEEE J. Biomed. Health Informatics* 27 (2) (2023) 866–877.
- [20] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, V. Chandra, Federated Learning with Non-IID Data, Cornell University, 2018, arXiv, Cornell University - arXiv.
- [21] Q. Li, Y. Diao, Q. Chen, B. He, Federated learning on non-IID data silos: An experimental study, in: 2022 IEEE 38th International Conference on Data Engineering, ICDE, 2022.
- [22] Z. Zhao, C. Feng, W. Hong, J. Jiang, C. Jia, T.Q.S. Quek, M. Peng, Federated learning with non-IID data in wireless networks, *IEEE Trans. Wirel. Commun.* (2022) 1927–1942.
- [23] L. Gao, H. Fu, L. Li, Y. Chen, M. Xu, C.-Z. Xu, 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, 2022, pp. 10112–10121.
- [24] Y. Sun, L. Shen, T. Huang, L. Ding, D. Tao, FedSpeed: Larger local interval, less communication round, and higher generalization accuracy, 2023, arXiv preprint arXiv:2302.10429.
- [25] Y. Sun, L. Shen, S. Chen, L. Ding, D. Tao, Dynamic regularized sharpness aware minimization in federated learning: Approaching global consistency and smooth landscape, in: International Conference on Machine Learning, PMLR, 2023, pp. 32991–33013.
- [26] Q. Pan, R. Wu, T. Liu, T. Zhang, Y. Zhu, W. Wang, Fedgkd: Unleashing the power of collaboration in federated graph neural networks, 2023, arXiv preprint arXiv:2309.09517.
- [27] T. Lin, L. Kong, S. Stich, M. Jaggi, Ensemble distillation for robust model fusion in federated learning, *Neural Inf. Process. Syst.*, Neural Inf. Process. Syst. (2020).
- [28] P. Ekman, W.V. Friesen, Facial action coding system, *Environ. Psychol. & Nonverbal Behav.* (1978).
- [29] K. Mase, Recognition of facial expression from optical flow, *IEICE Trans. Inf. Syst.* 74 (10) (1991) 3474–3483.
- [30] S. Li, W. Deng, Deep facial expression recognition: A survey, *IEEE Trans. Affect. Comput.* 13 (3) (2020) 1195–1215.
- [31] K. Wang, X. Peng, J. Yang, D. Meng, Y. Qiao, Region attention networks for pose and occlusion robust facial expression recognition, *IEEE Trans. Image Process.* 29 (2020) 4057–4069.
- [32] H. Wang, B. Li, S. Wu, S. Shen, F. Liu, S. Ding, A. Zhou, Rethinking the learning paradigm for dynamic facial expression recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 17958–17968.
- [33] Z. Sun, H. Zhang, J. Bai, M. Liu, Z. Hu, A discriminatively deep fusion approach with improved conditional gan (im-cGAN) for facial expression recognition, *Pattern Recognit.* 135 (2023) 109157.
- [34] H. Li, H. Niu, Z. Zhu, F. Zhao, Intensity-aware loss for dynamic facial expression recognition in the wild, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 37, 2023, pp. 67–75.
- [35] C. Liu, K. Hirota, Y. Dai, Patch attention convolutional vision transformer for facial expression recognition with occlusion, *Inform. Sci.* 619 (2023) 781–794.
- [36] T. Liu, J. Li, J. Wu, B. Du, J. Wan, J. Chang, Confusable facial expression recognition with geometry-aware conditional network, *Pattern Recognit.* 148 (2024) 110174.
- [37] J. He, Y. Xiao, H. Zhang, J. Cai, L. Cai, R. Liu, Micro_nest: multi-scale attention enhanced micro-expression recognition framework, *Expert Syst. Appl.* 290 (2025) 128372.
- [38] E. Barsoum, C. Zhang, C.C. Ferrer, Z. Zhang, Training deep networks for facial expression recognition with crowd-sourced label distribution, in: Proceedings of the 18th ACM International Conference on Multimodal Interaction, 2016, pp. 279–283.
- [39] S. Li, W. Deng, J. Du, Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2852–2861.
- [40] I.J. Goodfellow, D. Erhan, P.L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, et al., Challenges in representation learning: A report on three machine learning contests, in: *Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3–7, 2013. Proceedings, Part III* 20, Springer, 2013, pp. 117–124.
- [41] A. Mollahosseini, B. Hasani, M.H. Mahoor, Affectnet: A database for facial expression, valence, and arousal computing in the wild, *IEEE Trans. Affect. Comput.* 10 (1) (2017) 18–31.
- [42] B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in: *Artificial Intelligence and Statistics*, PMLR, 2017, pp. 1273–1282.
- [43] J. Xu, S. Wang, L. Wang, A.C.-C. Yao, Fedcm: Federated learning with client-level momentum, 2021, arXiv preprint arXiv:2106.10874.
- [44] R. Dai, X. Yang, Y. Sun, L. Shen, X. Tian, M. Wang, Y. Zhang, Fedgamma: Federated learning with global sharpness-aware minimization, *IEEE Trans. Neural Networks Learn. Syst.* (2023).
- [45] D. Caldarola, P. Cagnasso, B. Caputo, M. Ciccone, Beyond local sharpness: Communication-efficient global sharpness-aware minimization for federated learning, in: *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 25187–25197.

Yuanlun Xie received the bachelor's degree from Guizhou University, China, and the Ph.D. degree from the University of Electronic Science and Technology of China, China. He is now an assistant professor with the Chengdu University, China. His research interests include facial expression recognition, computer vision and deep learning. ORCID: 00000001-9682-2065.

Jinshan Lai Jinshan Lai graduated with a bachelor's degree from the University of Electronic Science and Technology of China (UESTC) and is currently pursuing a doctoral degree at UESTC. His research interests include federated learning, facial expression recognition, and computer vision. ORCID: 0000-0002-2222-9178.

Fengchun Zhang is a master's student at the University of Electronic Science and Technology of China, China. Her research interests include facial expression recognition and federated learning. ORCID: 0009-0001-7482-5932.

Kaibo Shi received his Ph.D. degree from the School of Automation Engineering at the University of Electronic Science and Technology of China. He is a professor of School of Electronic Information and Electrical Engineering, Chengdu University. From September 2014 to September 2015, he was a visiting scholar at the Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada. He was Research Assistant with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Taipa, from May 2016 to Jun 2016 and January 2017 to October 2017. He was also a Visiting Scholar with the Department of Electrical Engineering, Yeungnam University, Gyeongsan, Korea, from December 2019 to January 2020. His current research interests include stability theorem, robust control, sampled-data control systems, networked control systems, Lurie chaotic systems, power system and neural networks. He is the author or coauthor of over 150 research articles. He is a very active reviewer for many international journals. Dr. Shi was a recipient of the Highly Cited Researcher Award listed by Clarivate Analytics in 2021 and 2022, a 2022–2025 Most Cited Chinese Researcher. He serves as a Guest Editor of *Fractal and Fractional*, Guest Editor of the *Frontiers in Neurorobotics and Advanced Engineering Sciences*. ORCID: 0000-0002-9863-9229

Nan Zhou received his Ph.D. degree from Center for Robotics, University of Electronic Science and Technology of China, Chengdu, Sichuan. He was a joint Ph.D. student in University of Alberta, Canada, in 2015. He was a Postdoc Fellow at the Centre for Smart Health, Hong Kong Polytechnic University during 2020–2021. He was a Lecture with Chengdu University of Information Technology during 2017–2022. He is now an Associate Professor with Chengdu University, China. His research interests include machine learning, feature learning, computer vision and sparse modeling. ORCID: 0000-0002-0434-6231

Deepak Kumar Jain currently serves as an Associate Professor at the School of Artificial Intelligence and Control Science and Engineering at Dalian University of Technology in Dalian, China since July 2023. Before this, he held the position of Assistant Professor at Chongqing University of Posts and Telecommunications in Chongqing, China. Dr. Jain's educational journey began with a Bachelor of Engineering degree, which he earned in 2010 from Rajiv Gandhi Proudयोगiki Vishwavidyalaya in India. Subsequently, in 2012, he obtained a Master of Technology degree from Jaypee University of Engineering and Technology, also in India. Later, he completed his Ph.D. at the Institute of Automation, University of Chinese Academy of Sciences, Beijing, China in 2018. From 2014 to 2018, Dr. Jain was honored with the prestigious CAS-TWAS Presidential Fellowship. His expertise has also been acknowledged by the Shandong Taian Administration of Foreign Expert Affairs, designating him as a "Foreign Expert." Additionally, he has held positions as an Adjunct Professor at Symbiosis International University in Pune, India. Throughout his professional journey, Dr. Jain has made substantial contributions to multiple peer-reviewed conferences and has authored numerous research papers in prestigious scientific journals such as *Pattern Recognition Letters*, *Neurocomputing*, *Information Fusion*, and *Expert Systems with Applications*. He has approximately fifty publications in renowned journals, with a Citation Count of 3000. He held the role of lead guest editor in prestigious journals such as *Pattern Recognition Letters*, *Computer Communications*, *Multimedia Tools and Applications*, *Neural Computing and Applications*, and several others. Since 2023, he has been a member of the editorial board of the *Information Fusion* journal. Additionally, he serves as an area chair for the renowned conferences The 18th IEEE International Conference on Automatic Face and Gesture Recognition (FG) 2024 and 2025 and the IEEE International Joint Conference on Biometrics (IJCB) 2024. He has delivered keynote Talks in India, China, and Spain. Additionally, he served as the Principal Researcher in the project titled "The National Key Research and Development Program of China (2022YFE0204500) - Construction and Joint Research 'Belt and Road'".

Joint Laboratory on Industrial Internet of Things," funded by the Ministry of Science and Technology, China, with a budget of 4.99 million RMB. His research primarily revolves around deep learning, machine learning, pattern recognition, and computer vision. Furthermore, he holds the esteemed title of senior member of the Institute of Electrical and Electronics Engineers (IEEE). ORCID: 0000-0002-3400-1613

Vitomir Štruc is a Full Professor at the University of Ljubljana, Slovenia. His research interests include problems related to biometrics, computer vision, image processing, and machine learning. He (co-)authored more than 250 research papers for leading international peer reviewed journals and conferences in these and related areas. Vitomir is a Deputy Editor-in-Chief for the IEEE Transactions on Information Forensics and Security, a Subject Editor for Elsevier's Signal Processing and an Associate Editor-in-Chief for Pattern Recognition. He regularly serves on the organizing committees of visible international conferences. He was a Program Co-Chair for IJCB 2023, IEEE FG 2024, IJCB 2025 and WACV 2025, and currently acts as a General Co-Chair for WACV 2026 and Tutorial Co-Chair for CVPR 2026. Dr. Štruc is a Senior member of the IEEE, a member of IAPR, Slovenia's ambassador for the European Association for Biometrics (EAB) and the former president and current supervisory committee member of the Slovenian Pattern Recognition Society, the Slovenian member of IAPR. Vitomir was the VP Technical Activities for the IEEE Biometrics Council 2022–2025, and currently serves as the vice-chair of the IAPR Technical Committee on Biometrics (TC4) and a

member of the Supervisory Board of the EAB. He is also a member of the Steering Committee of the IEEE Face and Gesture and CVF/IEEE WACV conference series. ORCID: 0000-0002-3385-5780

Badong Chen received the Ph.D. degree in Computer Science and Technology from Tsinghua University, Beijing, China, in 2008. He is currently a professor with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China. His research interests are in signal processing, machine learning, artificial intelligence and robotics. He has authored or coauthored over 300 articles in various journals and conference proceedings, and has won the 2022 Outstanding Paper Award of IEEE Transactions on Cognitive and Developmental Systems. Dr. Chen serves (or has served) as an Associate Editor for several journals including IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Cognitive and Developmental Systems, IEEE Transactions on Circuits and Systems for Video Technology, Neural Networks and Journal of The Franklin Institute. He has served as a PC or SPC Member for prestigious conferences including UAI, IJCAI and AAAI, and served as a General Co-Chair of 2022 IEEE International Workshop on Machine Learning for Signal Processing. ORCID: 0000-00031710-3818