FED-PsyAU: Privacy-Preserving Micro-Expression Recognition via Psychological AU Coordination and Dynamic Facial Motion Modeling –Supplementary Material

Jingting Li[†1,2], Yu Qian[†1,3], Lin Zhao[1,2], Su-Jing Wang[*1,2]
[1]State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China
[2]Department of Psychology, University of the Chinese Academy of Sciences, Beijing, 100049, China
[3]School of Computer, Jiangsu University of Science and Technology, Zhenjiang, 212100, China
[†]Equal contribution, [*]Corresponding author (wangsujing@psych.ac.cn)

1. Introduction

This document provides additional details that could not be included in the main paper due to space constraints. In Sec. 3, we list all abbreviation in the main manuscript for reference, enhancing the readability. In Sec. 3, we present the statistical results of our psychological experiment. In Sec. 4, we explain the rationale behind selecting the 12 specific Action Units (AUs) used in the paper. Sec. 5 showcases the matrices representing intrinsic and extrinsic prior knowledge. Sec. 6 contains our full analysis of the confusion matrices. Sec. 7 describes the federal client's data situation. Sec. 8 includes supplementary results under the UAR metric in the context of federated learning. Additionally, Sec. 9 shows the ablation experiments we performed for different hyperparameter settings.

2. List of Abbreviations

Tab. 1 lists all abbreviations appearing in the text along with their corresponding full forms.

Table 1. List of Abbreviations in the main text

Abbreviation	Full Term
ME	Micro-expression
MER	Micro-expression Recognition
AU	Action Unit
OF	Optical Flow
FL	Federated Learning
ROI	Regions of Interest
GCN	Graph Convolutional Network
GAT	Graph Attention Network
LBP	Local Binary Pattern
MDMO	Main Directional Mean Optical Flow
MEGC	ME Grand Challenge
HDE	Holdout database Evaluation
CDE	Composite Database Evaluation
LRM	Localized ROI Modeling
LFE	Local ROI Feature Extractor
SSE	Spatial Structure Encoder
AFR	Relationship Modeling among AU Features
AFE	AU Feature Extractor
GSE	Group Squeeze and Excitation
FC	Fully Connected Layers
AP	AU Prediction
DPK-GAT	Dynamic Prior Knowledge GAT
DSI	Dual-Stream InceptionNet
FA	Federated Aggregation
P-FedProx	Personalized FedProx
LOSO	Leave-One-Subject-Out
UF1	Unweighted F1 Score
UAR	Unweighted Average Recall

3. More Details and Statistical Results on our Psychological Experiment

3.1. Rationale for the Number of Subjects

In psychological research, when sample data adheres to a normal distribution, more accurate statistical computations can be performed, leading to more reliable results. According to the Central Limit Theorem, as the sample size increases, the distribution of sample means approximates a normal distribution, even if the underlying data is not normally distributed. In psychological studies, it is generally accepted that a sample size of 30 or more participants is sufficiently large for the sample to approximate normality. Given that our study involves 30 participants, this sample

size is considered adequate for the purposes of the analysis.

3.2. Consistency in Emotional Expression across Cultures

Our study is based on the premise that emotional expression is universally consistent across cultures. Darwin, in his book The Expression of the Emotions in Man and Animals, argued that different facial expressions are innate and universal, understood by all humans. Contemporary research supports this view, such as Ekman's studies, where participants from 10 different countries and regions were shown 30 photos of faces expressing¥ six basic emotions (happiness, surprise, sadness, fear, disgust, and anger). The results demonstrated a high level of consistency in the recognition of these six emotions across cultures. Meantime, we acknowledge that incorporating a more diverse cultural background would enhance the reliability of our results, we plan to collaborate with other research institutions in the future to gather more data.

3.3. More Detail on AU-based Material Composition

The materials used in this study are facial images provided by Professor Ekman. The selected facial Action Units (AUs) represent one of the six basic emotions, including: AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU10, AU12, AU14, AU15, AU16, AU17, AU22, AU23, AU24, AU25, and AU43. During the image stitching process, we ensured that facial images with consistent expression intensity were used. The stitched images include both coordinating pairs (e.g., AU6 and AU12, AU2 and AU25) and mutually exclusive pairs (e.g., AU4 and AU12, AU6 and AU15).

3.4. Statistical Results

As described in the main paper, participants were tasked with evaluating each presented image combination to determine whether it represented a coordinated combination. They were also asked to rate the degree of coordination or lack thereof. Tab. 2 summarizes the participants' accuracy in judging the combinations (i.e., correctly identifying coordinated combinations as coordinated and uncoordinated combinations as uncoordinated) as well as their ratings for the degree of coordination or lack of coordination. Fig. 2 lists the AU combinations most frequently selected as coordinated or uncoordinated.

The results indicate that participants were more likely to accurately recognize coordinated AU combinations and assigned higher scores to their degree of coordination. This finding suggests that at the cognitive level, coordinated AUs—representing consistent emotional expressions across the upper and lower face—are more readily accepted and processed. In contrast, uncoordinated AU combinations conveyed atypical emotional expressions, which posed greater challenges for participants in recognizing the associated emotions.

Table 2. Descriptive statistics for the accuracy and ranking scores of coordinated AU combinations and mutually exclusive AU combinations. C-AU and N-AU represent the coordinated AU pairs and uncoordinated AU pairs, respectively. M and SD denote the mean value and the standard deviation.

	Accuracy rate		Score scale	
	C-AU	N-AU	C-AU	N-AU
M	0.813	0.544	109.633	54.8
SD	0.119	0.196	27.697	28.850

4. AU Selection Strategy

In this psychological experiment, we did not impose restrictions on the intensity of AU movements. All selected AUs are related to facial expressions, and our primary focus was on investigating the coordination or mutual exclusivity of different AU combinations. However, in the context of micro-expression (ME) expression, some AUs rarely appear. Therefore, for the MER algorithm design, we filtered the AUs based on their frequency of appearance in the ME database, retaining only the 12 AUs listed in Table 1 in the main paper.

In particular, We selected 12 AUs associated with ME features primarily based on statistical results from DFME, the largest current ME database, while also considering AU distributions in other databases, excluding those AUs that are not commonly observed across these sources. As shown Fig. 1, these AUs, i.e., AU 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, are chosen because they exhibit high activation strength and frequency. Notably, although AU20, AU23, AU24 and AU38 show higher activation levels than AU15 and AU9 in the DFME dataset, these three AUs were not significantly activated in CAS(ME)^3. In contrast, AU9 and AU15 exhibit distinguishable facial ME movements with notable activation levels across the two datasets. Since our goal for MER is to ensure validation across multiple datasets, it is crucial to select AUs that are consistently annotated and exhibit relatively high activation levels in all datasets. Therefore, AU9 and AU15 were chosen to replace AU23 and AU24. The emergence of these ME-related AUs has greatly helped MER.

5. Prior Knowledge

5.1. Intrinsic Prior Knowledge

The intrinsic prior Knowledge, rooted in a psychology study, forms the adjacency matrix of DPK-GAT. Specifically, in our psychological experiment, we primarily inves-

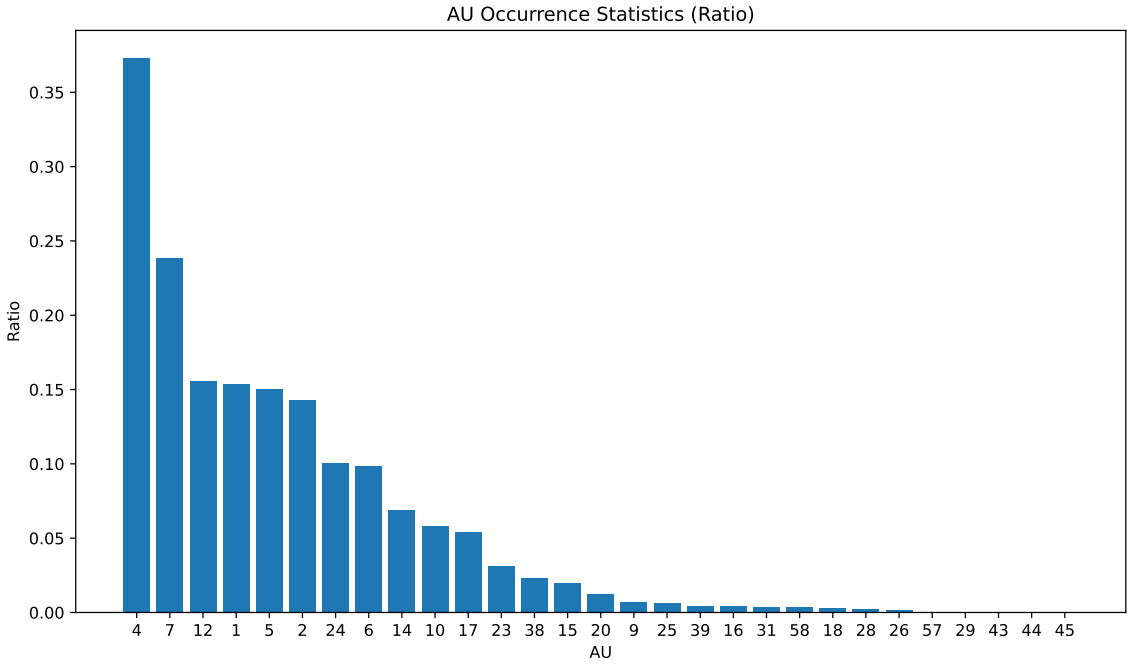


Figure 1. AU Occurrence associated with ME characteristics in DFME.

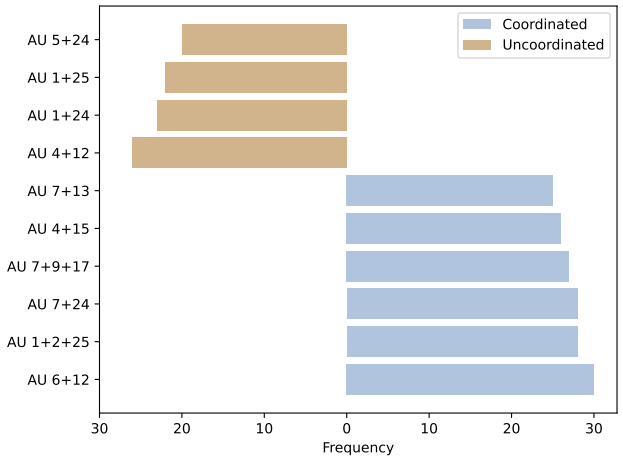


Figure 2. Statistics of the number of correct judgments for the coordinated AU combinations and mutually exclusive AU combinations. Combinations more than 20 times, i.e., exceeding two-thirds, are listed.

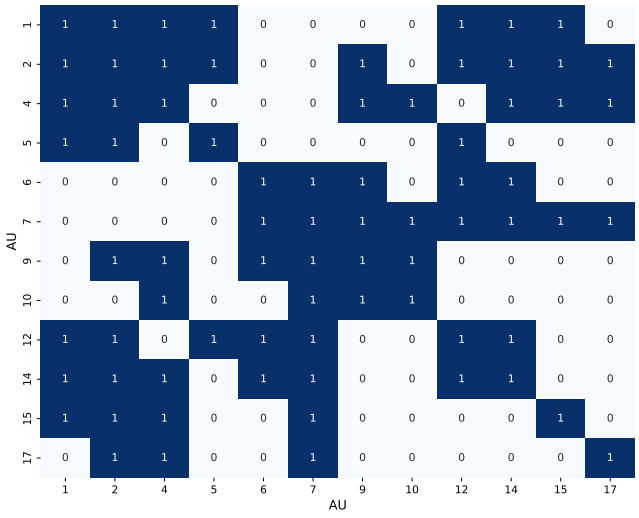


Figure 3. Intrinsic prior knowledge from psychology study.

tigated AU combinations that cannot be directly explained by physiological contradictions, specifically focusing on combinations where AUs are distributed across the upper and lower face, and their expressions are either coordinated or mutually exclusive. In the prior matrix for MER, we considered both aspects: one related to the physiological coordination or mutual exclusivity of AUs (e.g., AU4 frowning and AU5 widening the eyes, where the muscle movements corresponding to these two AUs cannot occur simultaneously), and the other related to the emotional expression and perception of AUs. We used both experimental results and prior knowledge about the muscle-emotion relationships to make further judgments. As shown in Fig. 3, this matrix illustrates the collaborative relationships between AUs, integrating anatomical physiology, emotional psychology, and our experimental findings. Specifically, the diagonal elements are always 1, indicating that each AU is entirely coordinated with itself. For each row, the off-diagonal ele-

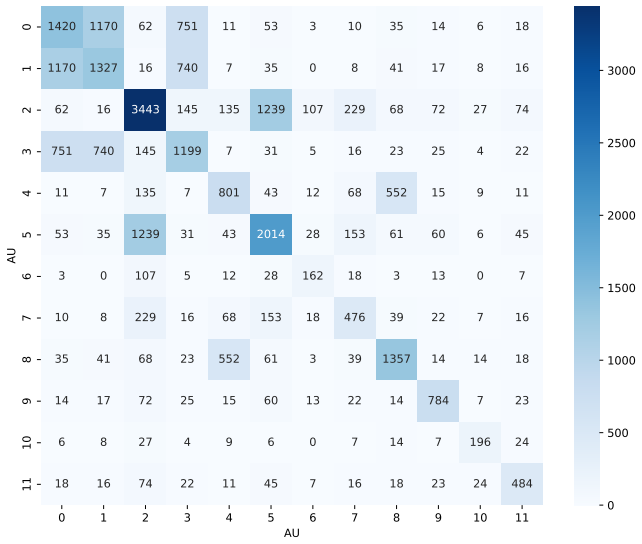


Figure 4. Extrinsic prior knowledge from statistical regularity in DFME.

ments with a value of 1 signify coordination with the diagonal element of that row. In contrast, those with a value of 0 indicate mutual exclusivity with the diagonal element. Additionally, it should be noted that our psychological study primarily demonstrated that perceptible mutual exclusions and coordinations exist between AUs in the upper and lower facial regions. However, the specific quantitative relationships between AUs are highly complex, varying with different action intensities and emotions. Therefore, we provided a simple binary (0-1) prior knowledge matrix to represent AU relationships and guide the network, rather than attempting a precise quantification.

5.2. Extrinsic Prior Knowledge

Using a sample of 7,526 MEs from the DFME dataset, we analyze the co-occurrence of AUs, as each ME sample is annotated with corresponding AUs or AU combinations. This analysis is further utilized to generate the prior attention matrix required by DPK-GAT. The data-driven co-occurrence matrix is called Extrinsic prior Knowledge and illustrated in Fig. 4.

6. Confusion Matrix

Here, we provide the confusion matrices further to discuss our model's performance on the MER task, as shown in Fig. 5. We found that our model achieved excellent results on categories of happiness and surprise in a seven-classification task on DFME, which suggests that our model further understands the facial topology by extracting the movement patterns of the facial muscles and improves the discriminative ability for different emotion categories. At the same time, we found that the model is prone to confounding when dealing with negative emotions; for example, the model often predicts anger as disgust and much of this result stems from the slight feature differentiation between samples of the same polarity emotions. The weak differences between these samples will likely be further masked by cropping and calculating the facial optical flow. For CAS(ME)^3, the model shows some degree of confounding in the positive categories due to too few samples in the positive categories of CAS(ME)^3 (negative: 457, positive: 55, surprise: 187), resulting in the model not learning enough discriminative representations between categories.

7. Federal Local Client Settings

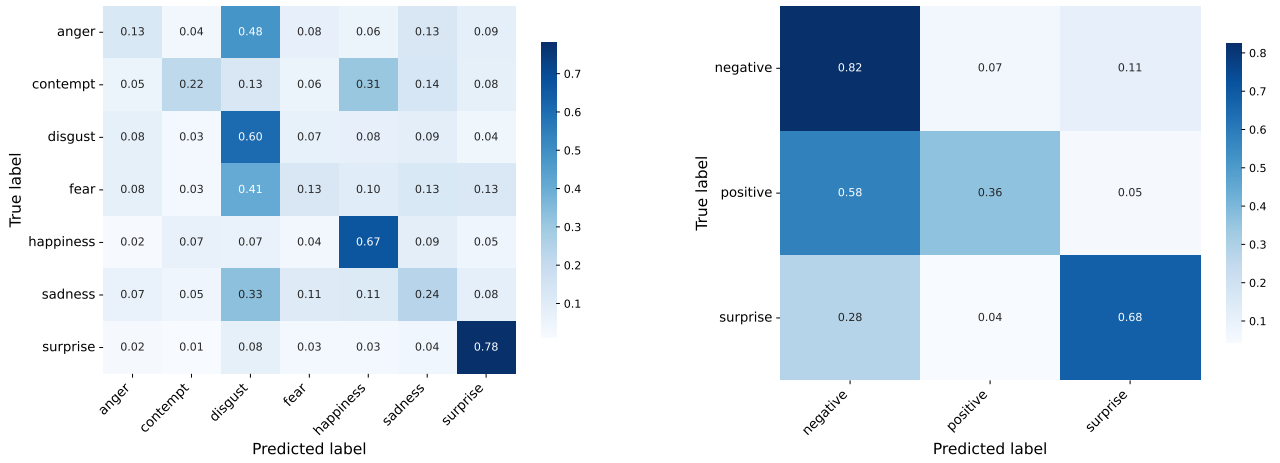
As we mentioned in the main paper, to simulate a realistic scenario in which data can not be shared, based on the number of subjects, the DFME and CAS(ME)^3 datasets are randomly partitioned into multiple local clients based on subject numbers. Specifically, DFME is divided into five equal parts, and CAS(ME)^3 into two equal parts. Notably, due to variations in the number of MEs per subject, the resulting clients, despite having an equal number of subjects, exhibit differences in ME data distribution and quantity. The specific amount of ME samples for each client is listed in Tab. 3.

Table 3. Federal Local Client Settings

Dataset	Client	ME Sample Size
DFME	Client1	1389
	Client2	1547
	Client3	1334
	Client4	1492
	Client5	1513
CAS(ME)^3	Client1	426
	Client2	273

8. Additional Federated Experimnet

As shown in Fig. 6, P-FedProx outperforms FedProx and FedAvg on two CAS(ME)^3-split local clients, demonstrating the value of client-specific global models. From our perspective, even though the data in DFME is divided among five local clients, the amount of data in these clients is still significantly larger than that in the clients from CAS(ME)^3. The disparities in both data size and distribution pose significant challenges to traditional federated learning methods, such as FedAvg and FedProx. When compared to recent personalized FL methods, P-FedProx achieves superior performance to FedRep and ELLP, and is competitive with FedAS, while requiring only a significantly more straightforward implementation.



(a) DFME (b) CAS(ME)^3

Figure 5. Confusion matrices for our model on the DFME and CAS(ME)^3 datasets.

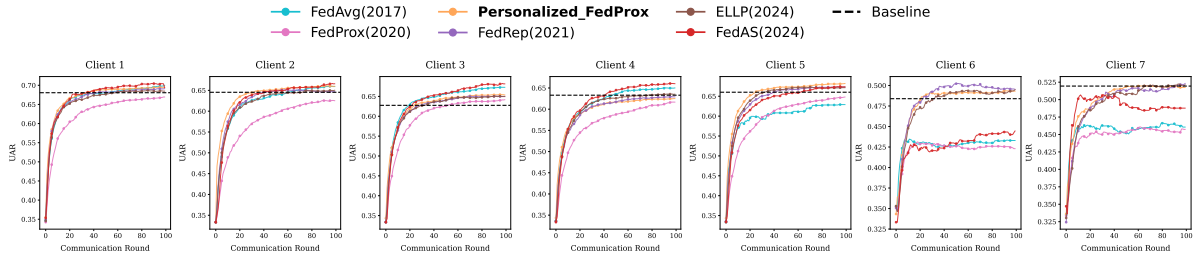


Figure 6. Compare the performance of different federated learning frameworks on different clients: UAR

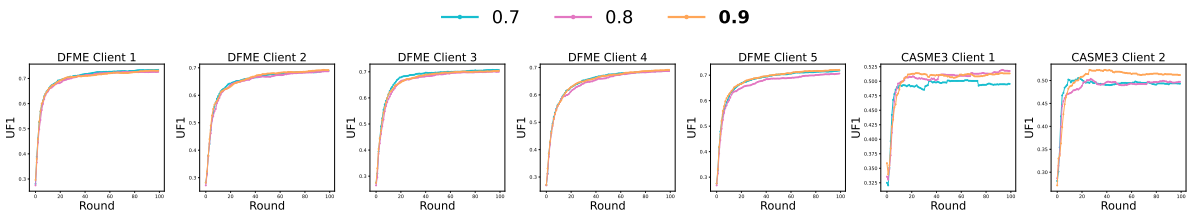
9. Hyperparameter Ablation Experiments

9.1. Model Fusion Weight for Local Client θ in Federated Learning

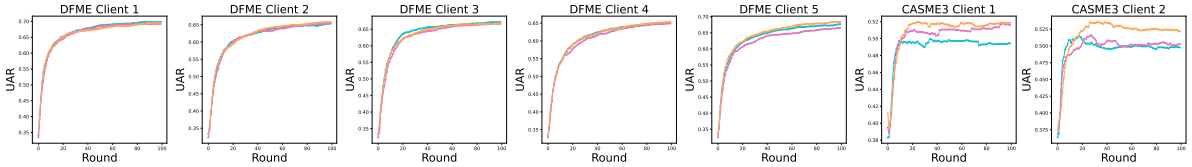
In our P-FedProx framework, θ controls the weight of the model fusion for each local client in the new communication round. Specifically, the larger the value of θ, the greater the influence of the locally trained model weight from the previous round on the new communication round, resulting in a stronger impact on the local training in the upcoming round. We conducted experiments on the effects of different θ (0.7,0.8,0.9), and the experimental results are shown in Fig. 7. We found that during the training process of most local clients, when θ is equal to 0.9, the model has better convergence and ability to cope with heterogeneous data. This means that in each federated communication round, smaller fusion weights for the remaining clients can help stabilize the local training of the current client, while enhancing feature extraction capability and reducing the impact of data heterogeneity on local training.

9.2. Loss Function Weight α

As written in the main paper, α is used to balance the weights of the components in the loss function. Specifically, α_1 is used to control the sentiment classification loss weights during model training, α_2 and α_3 are used to control the AU prediction loss weights, and α_4 is used to regulate the impact of the global model on local client training in each communication round of federated learning. Here, we provide experiments with different parameters α in the equations. 8 and 9, where α_1,α_2,α_3 in Eq. 8 are listed in Tab. 4 and α_4 in Eq. 9 are shown in Fig. 8. In the final evaluation of the model, we set α_1,α_2, and α_3 to 0.8, 0.2, and 0.2, respectively. We try to crank up the weight of MER loss to satisfy the final recognition needs while balancing the weight of AU recognition loss to guide the network in uncovering more accurate AU co-occurrence patterns. When further extended to the federated paradigm, as can be seen from the UF1 and UAR results in Fig. 8, P-FedProx achieves optimal results when α_4 is set to 0.001 in almost all local clients. We believe this is a balance between local training and global fusion. With more significant data heterogeneity, clients with less local data are more suscepti-

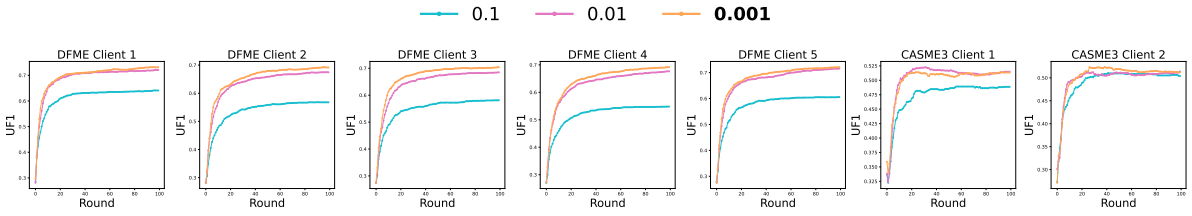


(a) The effect of different θ on UF1.

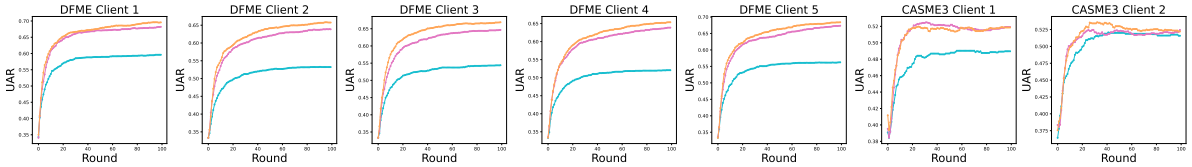


(b) The effect of different θ on UAR.

Figure 7. The effect of different θ on UF1 and UAR.



(a) The effect of different α_4 on UF1.



(b) The effect of different α_4 on UAR.

Figure 8. The effect of different α_4 on UF1 and UAR.

ble to the influence of other clients, so a more personalised training approach becomes effective.

Table 4. Comparison of UF1 and UAR for different values of α on the DFME dataset.

α_1	α_2	α_3	UF1	UAR
0.5	0.5	0.5	0.3701	0.3827
0.4	0.6	0.6	0.3698	0.3804
0.3	0.7	0.7	0.3715	0.3857
0.2	**0.8**	**0.8**	**0.3853**	**0.3978**
0.1	0.9	0.9	0.3744	0.3863

9.3. About Pre-training AU Moudle on DISFA and CK

To further improve the model's basic ability to perceive facial motion patterns, we first pre-trained the AU module (i.e., the part of the network prior to the DSI module) on the DISFA and CK datasets. The results showed that the pre-trained model significantly outperformed the non-pre-trained model, as shown in Tab. 5.

Table 5. Results on two datasets with and without the AU Group

Dataset	**Setting**	**UF1**	**UAR**
DFME	Without Pre-train	0.3761	0.3891
	With Pre-train	**0.3853**	**0.3978**
CAS(ME)^3	Without Pre-train	0.5983	0.5959
	With Pre-train	**0.6221**	**0.6226**