

Supplementary Material

Table 1. Implementation details of some comparison methods.

Methods	Details
RSCFed	$K=3, M=3$ for <i>ICH</i> and $K=6, M=10$ for <i>ChestXray14</i>
FixMatch	$\tau=0.8$
FedIRM	$T=20, \tau=2.0, \omega=30, ema=0.99$
CBAFed	$P=50, \alpha_1=0.8, \alpha_2=0.5, J=1, T=500, \tau=0.6$
FedLSR	$\lambda \sim Beta(1, 1), \gamma = Min[0.4 \cdot \frac{t}{40}, 0.4]$
FedNoRo	$\lambda_{max}=0.8$

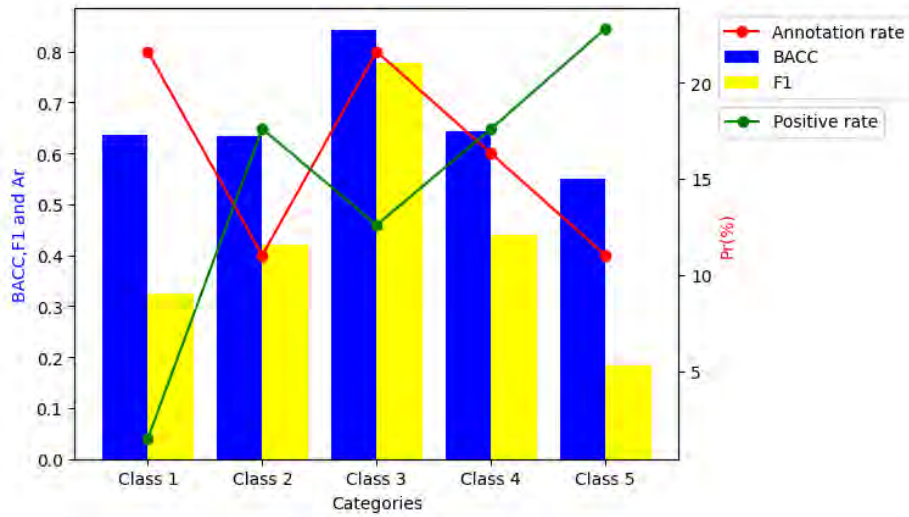


Fig. 1. Results of FedAvg on *ICH* where cool classes with lower annotation rates facing performance degradation. The recognition ability of class 1 is also weak due to the influence of intra-class imbalance.

Algorithm 1 FedMLP.

Input: Initialized global model θ_G^1 ; dataset D_k in client k amount N_k , $k \in [K]$; labeled class set AC_k and unlabeled class set NC_k in client k ; hyper-parameter L , R , T_0 and T_1 ; Warm-up rounds t_1 ; total communication rounds T ; local training epoch E .

```

0:  $\triangleright$  Stage1: Warm-up
1: for  $t \leftarrow 1$  to  $t_1$  do
2:   In local clients
3:   for  $k \leftarrow 1$  to  $K$  do
4:      $\theta_k^t \leftarrow \theta_G^t$ 
5:     for  $e \leftarrow 1$  to  $E$  do
6:       Augment for each sample in  $D_k$  as  $D_k^{Aug1}$  and  $D_k^{Aug2}$ 
7:        $\theta_k^t \leftarrow$  update by Eq. 4 with the local dataset  $D_k^{Aug1} \cup D_k^{Aug2}$ 
8:     end for
9:     if  $t = t_1$  then
10:      Local Calculation()
11:    end if
12:  end for
13:  In central server
14:  if  $t = t_1$  then
15:    Global Aggregation( $\theta_k^t, P_0^{k,c}, P_1^{k,c}, d_c^k$ )
16:  else
17:     $\theta_G^{t+1} \leftarrow \sum_{k=1}^K \frac{N_k}{\sum_{i=1}^K N_i} \theta_k^t$ 
18:  end if
19: end for
19:  $\triangleright$  Stage2: Missing Label Detection
20: for  $t \leftarrow t_1 + 1$  to  $T$  do
21:  In local clients
22:  for  $k \leftarrow 1$  to  $K$  do
23:    Download  $\tau_0, \tau_1, \theta_G^t, P_0^c, P_1^c$  and  $d_c^G$ 
24:    select samples and categories with pseudo-labels use Eq. 7 ~ 9
25:    for  $e \leftarrow 1$  to  $E$  do
26:       $\theta_k^t \leftarrow$  updated by Eq. 4 for hard labels and MSE for soft labels
27:    end for
28:    Local Calculation()
29:  end for
30:  In central server
31:  Global Aggregation( $\theta_k^t, P_0^{k,c}, P_1^{k,c}, d_c^k$ )
32: end for
33:
34: Global Aggregation( $\theta_k^t, P_0^{k,c}, P_1^{k,c}, d_c^k$ )
35:  $\theta_G^{t+1} \leftarrow \sum_{k=1}^K \frac{N_k}{\sum_{i=1}^K N_i} \theta_k^t$ 
36: Calculate  $P_0^c, P_1^c$  and  $d_c^G$  use Eq. 6 and 10
37: Calculate  $\tau_0$  and  $\tau_1$ 
38: Output:  $\theta_G^{t+1}, P_0^c, P_1^c, d_c^G, \tau_0$  and  $\tau_1$ .
39:
40: Local Calculation()
41: Calculate  $P_0^{k,c}$  and  $P_1^{k,c}$  for  $c \in AC_k$  use Eq. 5
42: Calculate  $d_c^k$  for  $c \in AC_k$  use Eq. 10
43: Output:  $P_0^{k,c}, P_1^{k,c}$  and  $d_c^k$ .
Output: Global model  $\theta_G^{T+1}$ .

```
