

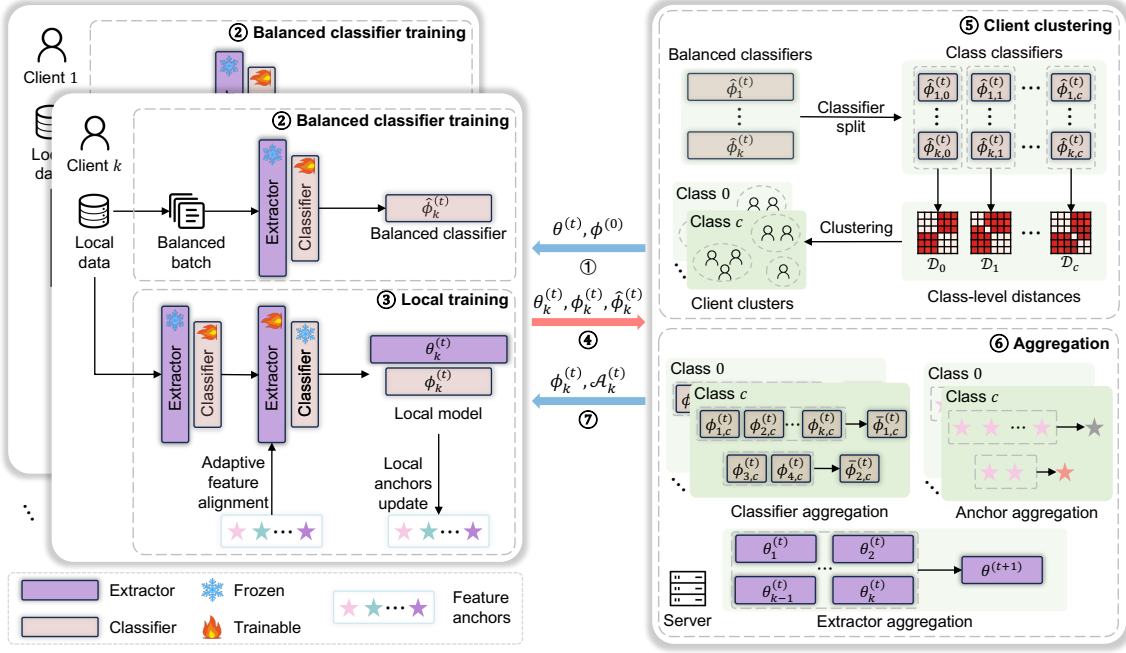


Figure 1: Overall pipeline of our proposed FedCCFA. At each round t , each selected client updates its extractor $\theta_k^{(t)}$ and classifier $\phi_k^{(t)}$ by $\theta^{(t)}$ and $\phi^{(0)}$, respectively. Then, client k performs balanced classifier training and local training, and sends its local parameters and feature anchors. After receiving all selected clients' parameters and feature anchors, the server performs client clustering, and then aggregates parameters and feature anchors according to the clustering results. Finally, each client updates its local classifier and feature anchors.

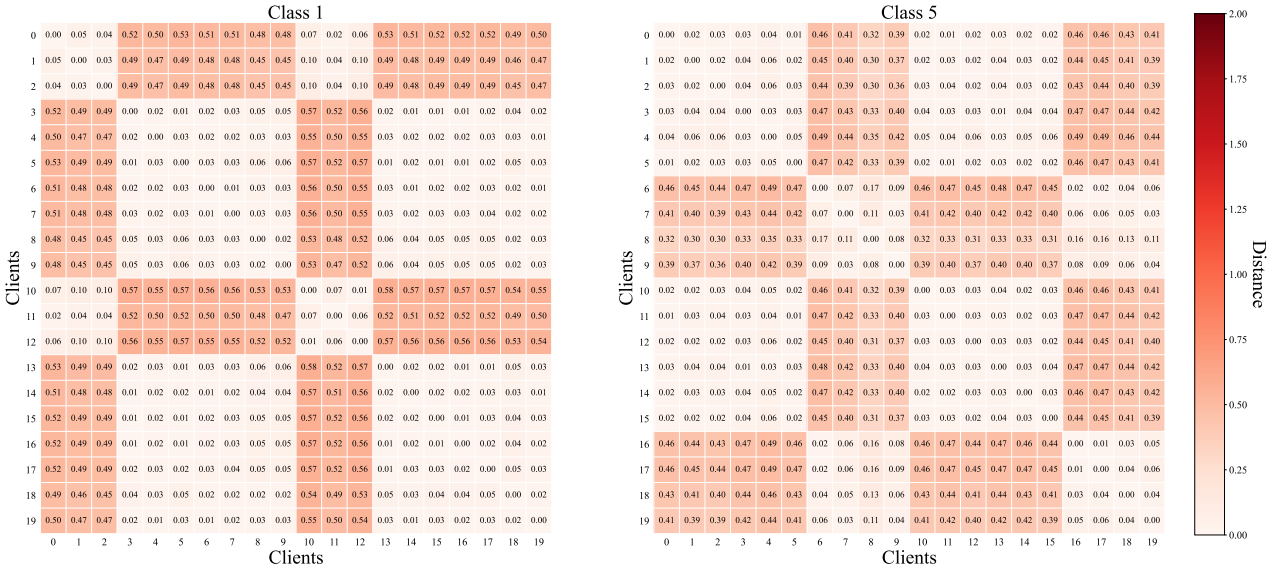


Figure 2: The visualization of distance matrix at round 199. We conduct experiments on CIFAR-10 under sudden drift setting. All 20 clients participate the training process. According to the concept drift settings in Section 5.1, for class 1, clients $\{0,1,2,10,11,12\}$ should be grouped and the others should be grouped; for class 5, clients $\{6,7,8,9,16,17,18,19\}$ should be grouped and the others should be grouped.