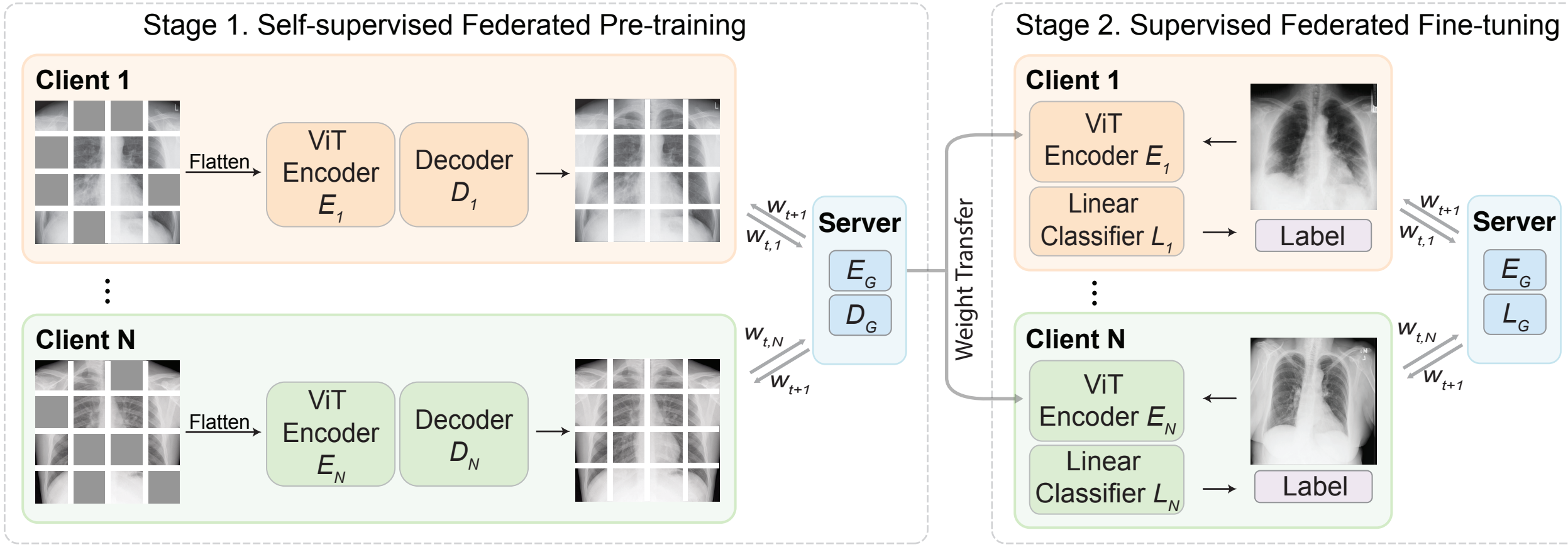
Stage 1. Self-supervised Federated Pre-training
Client 1
Flatten
ViT Encoder E_1
Decoder D_1
Client N
Flatten
ViT Encoder E_N
Decoder D_N
w_{t+1}
$w_{t,1}$
$w_{t,N}$
w_{t+1}
Server
E_G
D_G
Weight Transfer
Stage 2. Supervised Federated Fine-tuning
Client 1
ViT Encoder E_1
Linear Classifier L_1
Label
Client N
ViT Encoder E_N
Linear Classifier L_N
Label
w_{t+1}
$w_{t,1}$
$w_{t,N}$
w_{t+1}
Server
E_G
L_G