1 Bi-Level Meta-Learning Algorithm

To achieve the joint domain generalization (DG) of explanations and diagnoses, we design a bi-level meta-learning algorithm for training our GNN model with XG regularization. Initially, we randomly initialize the model parameters θ and gather fMRI data from multiple source-domain centers, i.e., $S = \{D_1, ..., D_K\}$, We then randomly partition S into inner-loop set S_{inner} and outer-loop set S_{outer}.

In the inner-loop of the bi-level optimization, we start by attaching initial values to our GNN model weights $\theta'_{k-1} = \theta_{k-1}$. we randomly sample a few subsets of source-domain centers (say S_{inner}), on which the GNN model is trained by simply L_{ce} for classification, this optimization process can be expressed as:

$$\theta'_k = \theta'_{k-1} - \gamma \times \frac{\partial L_{\text{ce}}(S_{\text{inner}}; \theta'_{k-1})}{\partial \theta'_{k-1}} = \theta_{k-1} - \gamma \nabla_{\theta_{k-1}} L_{\text{ce}}(S_{\text{inner}}; \theta_{k-1}), \tag{1}$$

where γ is step size of the Bi-Level Meta-Learning Algorithm.

In the outer-loop of the bi-level optimization, we further sample the remaining source-domain centers (say S_{outer}), on which the GNN model is trained by minimizing a meta-learning loss L_{meta}. We utilize the updated θ'_k to compute the loss function on S_{outer}:

$$\theta_k = \theta_{k-1} - \gamma \times \frac{\partial L_{\text{meta}}(S_{\text{outer}}; \theta'_k)}{\partial \theta_{k-1}}, \tag{2}$$

$$\frac{L_{\text{meta}}}{\partial \theta_{k-1}} = \frac{L_{\text{meta}}}{\partial \theta'_k} \frac{\partial \theta'_k}{\partial \theta_{k-1}}, \tag{3}$$

$$\frac{\partial \theta'_k}{\partial \theta_{k-1}} = \frac{\partial(\theta'_{k-1} - \gamma \times \frac{\partial L_{ce}}{\partial \theta'_{k-1}})}{\partial \theta_{k-1}} = 1 - \gamma \times \frac{\partial^2 L_{\text{ce}}}{\partial {\theta_{k-1}}^2}, \tag{4}$$

$$\theta_k = \theta_{k-1} - \gamma \times \frac{\partial L_{\text{meta}}}{\partial \theta'_k}(1 - \gamma \times \frac{\partial^2 L_{\text{ce}}}{\partial {\theta_{k-1}}^2}). \tag{5}$$

In order to prevent gradient explosion, higher-order fiducials $\frac{\partial^2 L_{\text{ce}}}{\partial {\theta_{k-1}}^2}$ is ignored, so the model parameters of the outer loop are updated with the equation:

$$\theta_k = \theta_{k-1} - \gamma \times \frac{\partial L_{\text{meta}}}{\partial \theta'_k} = \theta_{k-1} - \gamma \nabla_{\theta'_k} L_{\text{meta}}(S_{\text{outer}}; \theta'_k). \tag{6}$$