Legend: ⬤ set of ID data · ⬤(dashed) set of OOD data · — model trained with ✖-labeled U · ✖✖ ensemble predictions · model/ensemble predictions · ⬤⬤ labeled set S · ✖✖ artificially labeled U · — model trained with ✖-labeled U · ✖ identified as OOD in U · ensemble disagrees

(U_{OOD}, c) $(U_{\text{ID}}^{\neg c}, c)$ (U_{ID}^{c}, c)

a) Too little disagreement   b) Too much disagreement   c) Right amount of disagreement