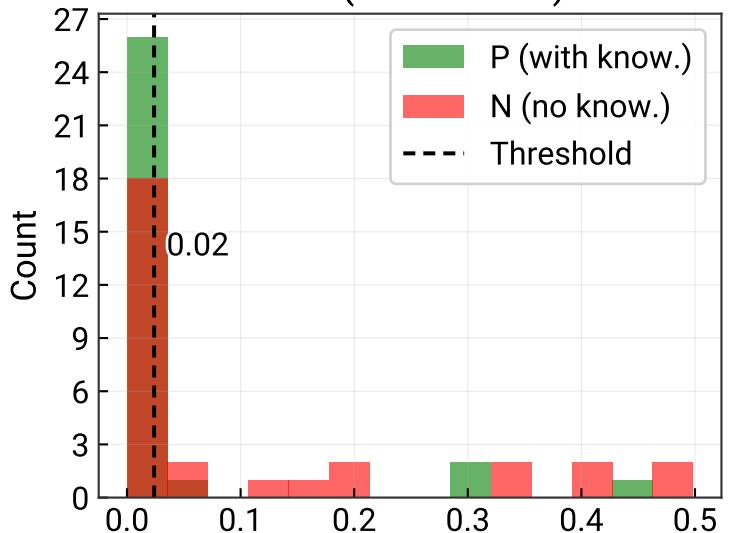
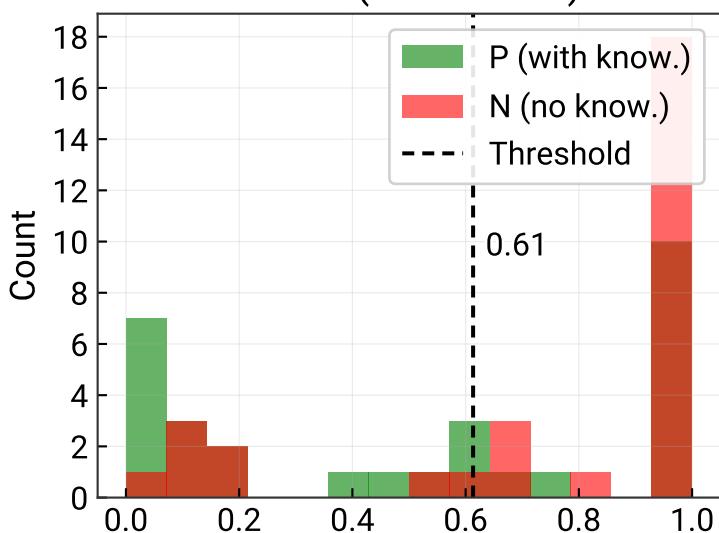


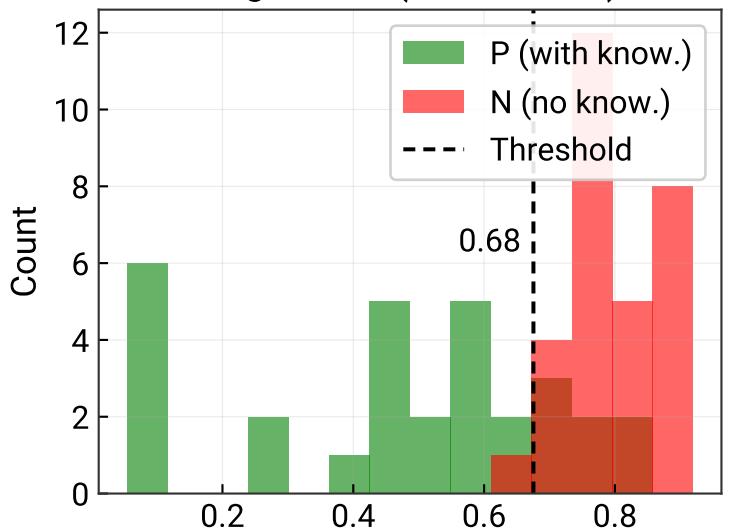
CKA (AUC: 0.648)



Fisher (AUC: 0.712)



Logit Lens (AUC: 0.927)



UDS (Ours) (AUC: 0.971)

