



Figure 2: Mean ranking of the methods across all seven sites by area under the receiver operating characteristic curve (AUROC). The error bars are 90% confidence intervals computed by the bootstrap. At left are the AUROC rankings from the leave-one-year-out cross validation (a), at right are the AUROC rankings from the leave-one-out cross validation (b).