Model benchmark using all features [10 repeats of 10–fold cross–validation]

C−Index (higher is better)

$\hat{\mu}_{mean} = 0.628$ (DeepSurv)
$\hat{\mu}_{mean} = 0.639$ (Ranger)
$\hat{\mu}_{mean} = 0.641$ (CTree)
$\hat{\mu}_{mean} = 0.645$ (CoxPH)
$\hat{\mu}_{mean} = 0.651$ (RF−SRC)
$\hat{\mu}_{mean} = 0.668$ (GBDT)
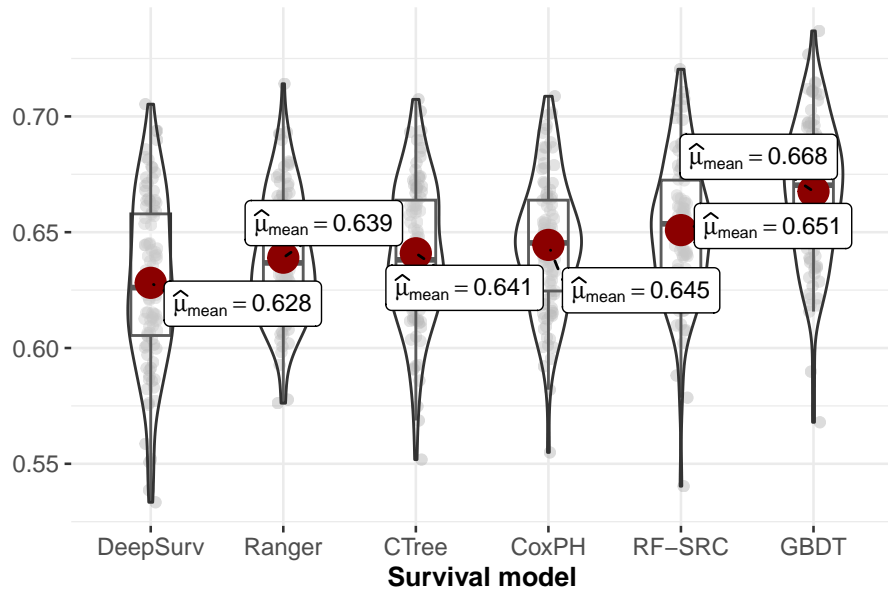
Integrated Brier Score (lower is better)

$\hat{\mu}_{mean} = 0.127$ (DeepSurv)
$\hat{\mu}_{mean} = 0.140$ (Ranger)
$\hat{\mu}_{mean} = 0.126$ (CTree)
$\hat{\mu}_{mean} = 0.127$ (CoxPH)
$\hat{\mu}_{mean} = 0.119$ (RF−SRC)
$\hat{\mu}_{mean} = 0.117$ (GBDT)

Survival model
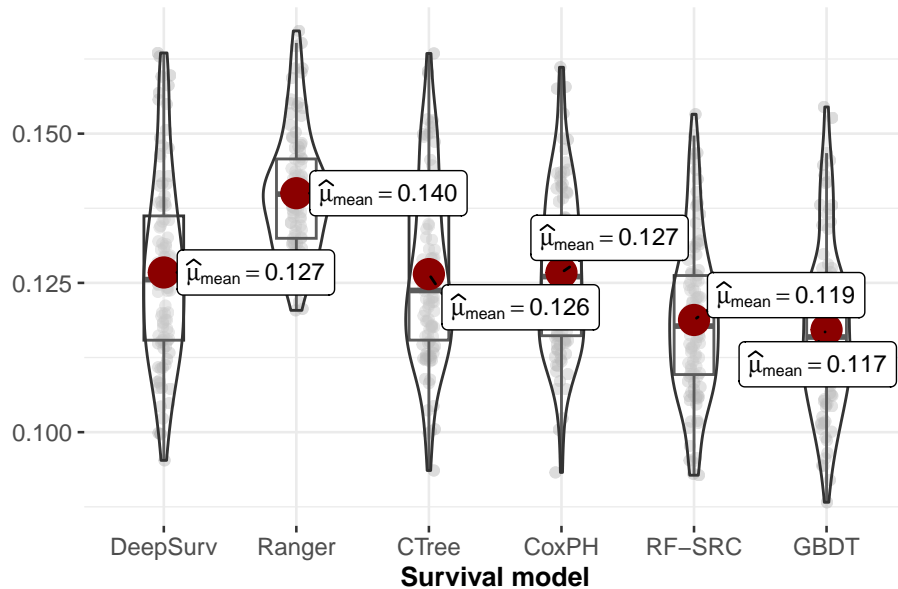
Note: Most of the DeepHit models did not converge and provide random predictions; thus, DeepHit is removed from the comparison.