Evaluation metrics: mean (by algorithm). Calibration dataset 1.00 -0.95 -0.90 -- GLM RF 0.85 -0.80 -0.75 -0.94 0.96 0.98 1.00 ROC