Evaluation metrics: boxplot (by algorithm, and PA). Validation dataset

