Testing Accuracy of different classifiers in different studies

Train/Test Split (%):

* Mouse All - 75/25
* Rat All - 75/25
* Rat Sterile - 60/40
* Rat PA - 65/35

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model / Study | Mouse All  (WT, B\_ENac, MPS I) | Rat All  (WT, KO, CF, Phe508) | Rat Sterile  (Baseline, Post-beads) | Rat PA  (WT, KO,CF) |
| SVM | 0.75 | 0.579 | 0.941 | 0.632 |
| Decision Tree | 0.75 | 0.526 | 0.824 | 0.737 |
| KNN | 0.688 | 0.684 | 0.882 | 0.842 |
| Random Forest | 0.625 | 0.579 | 0.824 | 0.632 |
| Gradient Boost | 0.75 | 0.684 | 0.824 | 0.474 |
| SGD | 0.75 | 0.526 | 0.941 | 0.526 |

Some of the datasets are too small, I increase the amount of testing data.

**Consider VDP, MSV, TV, VH, VHSS, VHLS, HD as features.**

**Different metrics (precision, recall (sensitivity) , F1 score, specificity, accuracy)**

**Train (mean, +- 10% deviation)//valid (mean, +- 10% deviation)/test,**

**Try different features**

**10-fold cross validation**