Analysis of the benefits of imputation models over traditional QSAR models for toxicity prediction

Moritz Walter1, Luke N Allen1, Antonio de la Vega de León1, Samuel J Webb2, Valerie J Gillet1

1 Information School, University of Sheffield, 211 Portobello, Sheffield S1 4DP, United Kingdom

2 Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds LS11 5PS, United Kingdom

**Table of content**

Table S1. Random Forest hyperparameters considered in optimization.

Table S2. XGBoost hyperparameters considered in optimization.

Table S3. Deep Neural Network hyperparameters considered in optimization.

Table S4. Macau hyperparameters considered in optimization.

Table S5. Median ROC-AUC values and interquartile ranges for assay-based splits.

Table S6. Selected auxiliary assays for TOX21-Aromatase-Inhibition

Figure S1. Performance of imputation models on the small datasets evaluated using ROC-AUC.

Figure S2. Performance of imputation models on the ToxCast dataset using ROC-AUC.

Figure S3. Performance of the imputation models using the GHOST approach

Figure S4. Ratio of Mutual Information (MI) and entropy of the target assay for Ames and Tox21 assays.

The tables S1-S4 provide the hyperparameters considered for optimization in a grid search. The optimal values found for each combination of split type, algorithm and assay can be found in the code repository (https://github.com/mowal/Imputation\_Paper).

**Table S1. Random Forest hyperparameters considered in optimization**.

For each hyperparameter modified in the grid search of Random Forest models, a brief explanation is provided alongside the values tested during the grid search.

|  |  |  |
| --- | --- | --- |
| **Hyperparameter** | **Explanation** | **Grid search values** |
| n\_estimators | Number of trees in the Random Forest model | 100, 500 |
| class\_weight | Assigns weights to instances of different classes | None, balanced, balanced\_subsample |
| max\_features | Number of randomly selected features that are considered to find the best split | sqrt, 0.25 |

**Table S2. XGBoost hyperparameters considered in optimization**.

For each hyperparameter modified in the grid search of XGBoost models, a brief explanation is provided alongside the values tested during the grid search.

|  |  |  |
| --- | --- | --- |
| **Hyperparameter** | **Explanation** | **Grid search values** |
| Num\_round | Number of sequential trees that are trained | 300, 700 |
| eta | Learning rate that shrinks the size of the update of predictions after each boosting step | 0.1, 0.3, 0.5 |
| Colsample\_bytree | Proportion of features used within each individual tree | 0.5, 0.7 |
| Alpha | Strength of L1 regularization of the weights per leaf | 0, 1 |
| lambda | Strength of L2 regularization of the weights per leaf | 1, 10 |
| Scale\_pos\_weight | Assign weights to instances of different classes | 1,weighted |

**Table S3. Deep Neural Network hyperparameters considered in optimization**.

For each hyperparameter modified in the grid search of Deep Neural Network models, a brief explanation is provided alongside the values tested during the grid search. ST: single task, MT: multi-task

|  |  |  |
| --- | --- | --- |
| **Hyperparameter** | **Explanation** | **Grid search values** |
| Hidden layers | Number of hidden layers in the network | 1, 2, 3  |
| Nodes per hidden layer | Number of nodes in each hidden layer | 1024, 2048 |
| Learning rate | Size of the updates to the network parameters per training step | 0.0003, 0.001, 0.003 |
| Dropout | Proportion of neurons randomly removed in each training step | 0, 0.2 |
| L2 regularization | Strength of L2 regularization on the weights per node | 0.0001, 0.001 |
| Batch size | Size of batches used for each update of weights with the Adam optimizer | 10, 50 |
| Number of epochs | Number of epochs (each data point is used once for training per epoch) | 3, 5, 10 |
| Class weight | Assigns weights to instances of different classes | 1, weighted [ST]1, 3, 5, 15 [MT] |

**Table S4. Macau hyperparameters considered in optimization**

For each hyperparameter modified in the grid search of Macau models, a brief explanation is provided alongside the values tested during the grid search.

|  |  |  |
| --- | --- | --- |
| **Hyperparameter** | **Explanation** | **Grid search values** |
| Num\_latent | Number of dimensions used for the latent space  | 16, 32, 64  |
| nsamples | Number of samples drawn from the Gibbs sampler | 800, 1600, 3200 |
| burnin | Number of burn-in samples from the Gibbs sampler that are discarded | 200, 400, 800 |

**Table S5 Median ROC-AUC values and interquartile ranges for assay-based splits.** Single task models are included as a benchmark. Median scores and interquartile ranges for each technique and dataset across 20 different random seeds. Before computing the median, the mean across the different assays for a single run was calculated. The best model for each dataset is in bold.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Ames | Tox21 |
| Single task | RF | 0.848 (0.846-0.849) | 0.817 (0.815-0.818) |
| XGB | 0.848 (0.846-0.848) | 0.805 (0.804-0.807) |
| ST-DNN | 0.831 (0.829-0.833) | 0.780 (0.778-0.783) |
| Multi-task | XGB-FN | 0.923 (0.923-0.924) | 0.866 (0.865-0.867) |
| MT-DNN | 0.935 (0.934-0.938) | 0.867 (0.864-0.870) |
| Macau | **0.944** (0.943-0.945) | **0.888** (0.887-0.888) |

**Table S6 Selected auxiliary assays for TOX21-Aromatase-Inhibition.** Assays were either selected according to their MI-entropy ratio or randomly.

|  |  |  |  |
| --- | --- | --- | --- |
| **Assays selected according to MI-entropy ratio** | **MI-entropy ratio** | **Randomly selected assays** | **MI-entropy ratio** |
| TOX21\_TR\_LUC\_GH3\_Antagonist | 0.369 | ATG\_RARb\_TRANS\_dn | 0.018 |
| TOX21\_AR\_BLA\_Antagonist\_ratio | 0.308 | TOX21\_MMP\_ratio\_up | 0.019 |
| BSK\_SAg\_CD40\_down | 0.295 | ATG\_Oct\_MLP\_CIS\_up | 0.124 |
| BSK\_SAg\_CD69\_down | 0.289 | ATG\_DR4\_LXR\_CIS\_dn | 0.136 |
| TOX21\_AR\_LUC\_MDAKB2\_Antagonist2 | 0.287 | ATG\_HSE\_CIS\_dn | 0.003 |
| TOX21\_AR\_LUC\_MDAKB2\_Antagonist | 0.284 | OT\_ER\_ERaERa\_1440 | 0.007 |
| BSK\_LPS\_SRB\_down | 0.279 | ATG\_Ets\_CIS\_dn | 0.048 |
| BSK\_4H\_Pselectin\_down | 0.276 | BSK\_3C\_HLADR\_down | 0.253 |
| BSK\_3C\_Proliferation\_down | 0.270 | BSK\_KF3CT\_IP10\_down | 0.209 |
| BSK\_3C\_Vis\_down | 0.270 | BSK\_hDFCGF\_TIMP1\_down | 0.206 |
| NCCT\_HEK293T\_CellTiterGLO | 0.269 | TOX21\_HSE\_BLA\_agonist\_ch2 | 0.043 |
| BSK\_SAg\_SRB\_down | 0.267 | TOX21\_VDR\_BLA\_agonist\_ch2 | 0.004 |
| BSK\_3C\_SRB\_down | 0.267 | NVS\_ADME\_hCYP19A1 | 0.073 |
| TOX21\_FXR\_BLA\_antagonist\_ratio | 0.267 | APR\_HepG2\_MicrotubuleCSK\_24h\_dn | 0.047 |
| BSK\_CASM3C\_Proliferation\_down | 0.265 | BSK\_SAg\_MIG\_down | 0.246 |
| BSK\_SAg\_CD38\_dow | 0.261 | BSK\_LPS\_MCP1\_down | 0.261 |
| BSK\_LPS\_MCP1\_dow | 0.261 | ATG\_NRF2\_ARE\_CIS\_up | 0.121 |
| BSK\_3C\_IL8\_down | 0.261 | BSK\_3C\_uPAR\_down | 0.214 |
| BSK\_SAg\_IL8\_down | 0.257 | NHEERL\_ZF\_144hpf\_TERATOSCORE\_up | 0.131 |
| BSK\_SAg\_MCP1\_down | 0.256 | TOX21\_PPARd\_BLA\_Antagonist\_ch1 | 0.012 |



**Figure S1** **Performance of imputation models on the small datasets evaluated using ROC-AUC.** A: Ames dataset. B: Tox21 dataset. Each box summarizes the ROC-AUC scores of 20 independent runs of the model on the test set with identical hyperparameters but different random seeds. The XGB models (best performing single task QSAR model based on MCC) are included as a benchmark. Only the best performing Feature Net model (XGB-FN) is included in this plot



**Figure S2 Performance of imputation models on the ToxCast dataset using ROC-AUC.** ROC-AUC scores obtained for the individual assays are sorted in descending order and plotted as a line.



**Figure S3 Performance of the imputation models using the GHOST approach.** Scatter plots contrasting MCC scores with or without the GHOST approach for single assays of the ToxCast dataset. A: XGB, B: XGB-FN, C: multi-task-DNN. The scatter plot for Macau is depicted in Figure 7B.



**Figure S4. Ratio of Mutual Information (MI) and entropy of the target assay for Ames (A) and Tox21 (B) assays.**