# Flame model documentation

# General model information

|  |  |
| --- | --- |
| ID | BSEP |
| Version |  |
| Model title | Support vector machine model |
| Model description | The model predicts whether a small molecule will inhibit the BSEP or not. |
| Keywords |  |
| Contact | Karolina Kopanska (karolinaweronika.kopanska@upf.edu) |
| Institution | Universitat Pompeu Fabra (UPF) |
| Date | May 2021 |
| Endpoint | Inhibition of the hepatic bile-salt export pump (BESP) |
| Endpoint units |  |
| Interpretation | A positive result indicates a high probability of being an inhibitor while a negative result suggests no inhibitory properties. |
| Dependent variable |  |
| Species |  |
| Limits applicability |  |
| Experimental protocol |  |
| Model availability | |  |  | | --- | --- | | description |  | | license |  | | location |  | |
| Data info | |  |  | | --- | --- | | endpoint negative |  | | endpoint positive |  | | format | SDF | | pre-procesing | Data reporting IC50 values for ABCB11 inhibition were extracted from 3 publications Warner et al., 2012 (DOI 10.1124/dmd.112.047068), Dawson et al., 2012 (DOI 10.1124/dmd.111.040758) and Morgan et al., 2010 (DOI 10.1093/toxsci/kfq269). Further data was gathered from open-source databases ChEMBL and PubChem. The ChEMBL data was filtered according to the following criteria Assay type "B" | "F"; Bioactivity type "IC50" | "Ki"; Activity operator "=". The PubChem data was filtered by Bioactivity type "IC50" | "Ki". Compounds without annotated activity values were excluded. All retrieved chemical compounds were standardized according to the Atkinson standardization protocol. Bonds to alkali metals and alkaline earth metals are removed. The entries then get checked for multiple fragments and each fragment is standardized separately. Non-organic compounds are removed, functional groups normalized, neutralized and finally stereochemistry is removed. Compounds were labeled as an inhibitor or non-inhibitor, if activity values were below 10 Î¼M or above 50 Î¼M, respectively. Compounds with activity values between 10 â€“ 50 Î¼M were removed. Duplicate compounds between the different sources were excluded if the classification labels disagreed. | | raw data url |  | | test selection | Random sample from training set | | test set size |  | | test set url |  | | training set size | 316 | | training set url | https://synapsemanagers.sharepoint.com/:u:/r/sites/etransafe/wp8/Shared%20Documents/Data%20and%20models%20repository/000019/Training.sdf?csf=1&web=1&e=Un93cY | |

# Algorithm and software

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm | |  |  | | --- | --- | | algorithm | SVM | | descriptors | ['RDKit\_md'] | | type | QSAR | |
| Software | |  |  | | --- | --- | | applicability domain |  | | descriptors | RDKit | | model | Flame version 1.0.0-rc2 | |
| Descriptors | |  |  | | --- | --- | | descriptors |  | | final number | 207 | | initial number |  | | ratio | 0.66 | | selection |  | | selection method |  | |
| Algorithm settings | |  |  | | --- | --- | | C | 10.0 | | break ties | False | | cache size | 200 | | class weight | None | | coef0 | 0.0 | | decision function shape | ovr | | degree | 1 | | gamma | auto | | kernel | rbf | | max iter | -1 | | probability | True | | random state | 42 | | shrinking | True | | tol | 0.001 | | verbose | False | |
| AD method | |  |  | | --- | --- | | description | Adding conformal features to the classifier model guarantees both, a certain level of confidence and a maximum frequency of errors which will be committed by the conformal predictor. Conformal accuracy indicates the fraction of correct predictions at the given confidence level. In classifier problems, the prediction region can include no class, one class or both classes. | | name | conformal prediction | |
| AD parameters | |  |  | | --- | --- | | confidence | 0.8 | | ACP sampler | BootstrapSampler | | KNN NN | 15 | | aggregated | True | | aggregation function | median | | conformal predictors | 10 | | normalizing model | KNN | |
| Goodness of fit statistics | |  |  | | --- | --- | | Specificity | 1.0 | | MCC | 0.98 | |
| Internal validation 1 | |  |  | | --- | --- | | Conformal coverage | 0.84 | | Conformal accuracy | 0.99 | |
| Internal validation 2 |  |
| External validation |  |
| Comments |  |

# Other information

|  |  |
| --- | --- |
| Other related models |  |
| Date of QMRF |  |
| Date of QMRF updates |  |
| QMRF updates |  |
| References | DOI 10.1124/dmd.112.047068; DOI 10.1093/toxsci/kfq269; DOI 10.1124/dmd.111.040758 |
| QMRF same models |  |
| Mechanistic basis |  |
| Mechanistic references |  |
| Supporting information |  |
| Comment on the endpoint | ABCB11 encodes the bile salt export pump (BSEP) which is expressed at the apical membrane of the hepatocytes. BSEP is responsible for the export of endogenous substrates and bile salts to the bile. Inhibition of the bile salt export pump (BESP) is considered an important mechanism leading to drug-induced liver injury (DILI). The hepatotoxic effect is triggered as a consequence of bile salt retention which may further lead to severe cholestasis and liver damage. |
| Endpoint data quality and variability |  |
| Descriptor selection |  |