

General model information (FHGGKU7K16)

[Edit](#)

	Value	Description																												
ID	BZR	Model identifier																												
Version	Development 27	Version of the model																												
Contact	Karolina Kopanska (karolinaweronika.kopanska@upf.edu)	Model identifier																												
Institution	Universitat Pompeu Fabra (UPF), Research Program on Biomedical Informatics (GRIB), Pharmacoinformatics group (Phi)	Affiliation details																												
Date	22.07.2020	Date of model development																												
Endpoint	Allosteric modulation of benzodiazepine receptor	Endpoint modeled																												
Endpoint units	Class prediction (active or inactive)	Response units to model																												
Interpretation	Benzodiazepine receptors are allosteric modulatory sites on GABAA receptors. The effect of the benzodiazepine receptor modulation is the potentiation of the neural inhibition that is mediated by gamma-aminobutyric acid (GABA). The BZR model classifies substances as active or inactive, thereby predicting their physiological effect on the receptor. A positive result indicates that the particular substance was classified as BZR modulator.	Interpretation of the model																												
Dependent variable	Activity class membership	Response variable used to model																												
Species		Modeled endpoint species																												
Limits applicability		Applicability limits of the model																												
Experimental protocol		Description of the followed experimental protocol																												
Model availability	<table border="1"> <tr> <td>description</td> <td>Additional information</td> </tr> <tr> <td>license</td> <td>Model license specifications</td> </tr> <tr> <td>location</td> <td>Model URL/place</td> </tr> </table>	description	Additional information	license	Model license specifications	location	Model URL/place	Information about model availability																						
description	Additional information																													
license	Model license specifications																													
location	Model URL/place																													
Data info	<table border="1"> <tr> <td>format</td> <td>SDF</td> <td>Format used (SDF, SMILES, table..)</td> <td>Model data resources</td> </tr> <tr> <td>pre-processing</td> <td>Mol blocks were computed from SMILES using RDKit.Chem</td> <td>Details on data processing</td> <td></td> </tr> <tr> <td>raw data url</td> <td>https://pubs.acs.org/doi/suppl/10.1021/ci5001168/suppl_file/ci5001168_sl_001.xlsx</td> <td>URL of the original data</td> <td></td> </tr> <tr> <td>test set size</td> <td></td> <td>Size of the test set</td> <td></td> </tr> <tr> <td>test set url</td> <td></td> <td>Test set URL</td> <td></td> </tr> <tr> <td>training set size</td> <td>130</td> <td>Size of the training set</td> <td></td> </tr> <tr> <td>training set url</td> <td></td> <td>Training set URL</td> <td></td> </tr> </table>	format	SDF	Format used (SDF, SMILES, table..)	Model data resources	pre-processing	Mol blocks were computed from SMILES using RDKit.Chem	Details on data processing		raw data url	https://pubs.acs.org/doi/suppl/10.1021/ci5001168/suppl_file/ci5001168_sl_001.xlsx	URL of the original data		test set size		Size of the test set		test set url		Test set URL		training set size	130	Size of the training set		training set url		Training set URL		
format	SDF	Format used (SDF, SMILES, table..)	Model data resources																											
pre-processing	Mol blocks were computed from SMILES using RDKit.Chem	Details on data processing																												
raw data url	https://pubs.acs.org/doi/suppl/10.1021/ci5001168/suppl_file/ci5001168_sl_001.xlsx	URL of the original data																												
test set size		Size of the test set																												
test set url		Test set URL																												
training set size	130	Size of the training set																												
training set url		Training set URL																												

Algorithm and software

[Edit](#)

	Value	Description																																																												
Algorithm	<table border="1"> <tr> <td>algorithm</td> <td>SVM</td> <td>Type of algorithm (Random Forest, SVM, etc)</td> <td>Description of the algorithm</td> </tr> <tr> <td>descriptors</td> <td>RDKit_properties</td> <td>Name of descriptor package/s used</td> <td></td> </tr> <tr> <td>type</td> <td>QSAR</td> <td>Type of model (QSAR, RA, etc)</td> <td></td> </tr> </table>	algorithm	SVM	Type of algorithm (Random Forest, SVM, etc)	Description of the algorithm	descriptors	RDKit_properties	Name of descriptor package/s used		type	QSAR	Type of model (QSAR, RA, etc)																																																		
algorithm	SVM	Type of algorithm (Random Forest, SVM, etc)	Description of the algorithm																																																											
descriptors	RDKit_properties	Name of descriptor package/s used																																																												
type	QSAR	Type of model (QSAR, RA, etc)																																																												
Software	<table border="1"> <tr> <td>applicability domain</td> <td></td> <td>Applicability domain software, version, description and license.</td> <td>Description of the used software</td> </tr> <tr> <td>descriptors</td> <td>RDKit</td> <td>Software used to generate descriptors, version, description and license</td> <td></td> </tr> <tr> <td>model</td> <td>Flame</td> <td>Main modelling program, version, description and license.</td> <td></td> </tr> </table>	applicability domain		Applicability domain software, version, description and license.	Description of the used software	descriptors	RDKit	Software used to generate descriptors, version, description and license		model	Flame	Main modelling program, version, description and license.																																																		
applicability domain		Applicability domain software, version, description and license.	Description of the used software																																																											
descriptors	RDKit	Software used to generate descriptors, version, description and license																																																												
model	Flame	Main modelling program, version, description and license.																																																												
Descriptors	<table border="1"> <tr> <td>descriptors</td> <td></td> <td>Descriptors used, list</td> <td>Information on descriptors used for the model</td> </tr> <tr> <td>final number</td> <td>25</td> <td>Final number of descriptors after feature selection</td> <td></td> </tr> <tr> <td>initial number</td> <td></td> <td>Initial number of descriptors</td> <td></td> </tr> <tr> <td>ratio</td> <td>0.19</td> <td>Ratio samples/descriptors</td> <td></td> </tr> <tr> <td>selection</td> <td></td> <td>Final list of descriptors used</td> <td></td> </tr> <tr> <td>selection method</td> <td></td> <td>Feature selection technique</td> <td></td> </tr> </table>	descriptors		Descriptors used, list	Information on descriptors used for the model	final number	25	Final number of descriptors after feature selection		initial number		Initial number of descriptors		ratio	0.19	Ratio samples/descriptors		selection		Final list of descriptors used		selection method		Feature selection technique																																						
descriptors		Descriptors used, list	Information on descriptors used for the model																																																											
final number	25	Final number of descriptors after feature selection																																																												
initial number		Initial number of descriptors																																																												
ratio	0.19	Ratio samples/descriptors																																																												
selection		Final list of descriptors used																																																												
selection method		Feature selection technique																																																												
Algorithm settings	<table border="1"> <tr> <td>C</td> <td>1</td> <td>na</td> <td>Algorithm parameters information</td> </tr> <tr> <td>break ties</td> <td>false</td> <td>na</td> <td></td> </tr> <tr> <td>cache size</td> <td>200</td> <td>na</td> <td></td> </tr> <tr> <td>class weight</td> <td></td> <td>na</td> <td></td> </tr> <tr> <td>coef0</td> <td>0</td> <td>na</td> <td></td> </tr> <tr> <td>decision function shape</td> <td>ovr</td> <td>na</td> <td></td> </tr> <tr> <td>degree</td> <td>3</td> <td>na</td> <td></td> </tr> <tr> <td>gamma</td> <td>auto</td> <td>na</td> <td></td> </tr> <tr> <td>kernel</td> <td>rbf</td> <td>na</td> <td></td> </tr> <tr> <td>max iter</td> <td>-1</td> <td>na</td> <td></td> </tr> <tr> <td>probability</td> <td>true</td> <td>na</td> <td></td> </tr> <tr> <td>random state</td> <td>46</td> <td>na</td> <td></td> </tr> <tr> <td>shrinking</td> <td>true</td> <td>na</td> <td></td> </tr> <tr> <td>tol</td> <td>0.001</td> <td>na</td> <td></td> </tr> <tr> <td>verbose</td> <td>false</td> <td>na</td> <td></td> </tr> </table>	C	1	na	Algorithm parameters information	break ties	false	na		cache size	200	na		class weight		na		coef0	0	na		decision function shape	ovr	na		degree	3	na		gamma	auto	na		kernel	rbf	na		max iter	-1	na		probability	true	na		random state	46	na		shrinking	true	na		tol	0.001	na		verbose	false	na		
C	1	na	Algorithm parameters information																																																											
break ties	false	na																																																												
cache size	200	na																																																												
class weight		na																																																												
coef0	0	na																																																												
decision function shape	ovr	na																																																												
degree	3	na																																																												
gamma	auto	na																																																												
kernel	rbf	na																																																												
max iter	-1	na																																																												
probability	true	na																																																												
random state	46	na																																																												
shrinking	true	na																																																												
tol	0.001	na																																																												
verbose	false	na																																																												
AD method	<table border="1"> <tr> <td>description</td> <td>Adding conformal features to the classifier guarantees both, a certain level of confidence and a maximum frequency of errors (misclassifications) which will be committed by the conformal predictor. The given confidence level opposes the error rate which represents the level of uncertainty that must initially be accepted and provided by the user. Conformal accuracy indicates the fraction of correct predictions at the given confidence level. Conformal coverage is the fraction of substances for which valid predictions were generated at the required confidence level.</td> <td>Description of the applicability domain technique</td> <td>Applicability domain technique details</td> </tr> <tr> <td>name</td> <td>conformal prediction</td> <td>Name of the applicability domain technique</td> <td></td> </tr> </table>	description	Adding conformal features to the classifier guarantees both, a certain level of confidence and a maximum frequency of errors (misclassifications) which will be committed by the conformal predictor. The given confidence level opposes the error rate which represents the level of uncertainty that must initially be accepted and provided by the user. Conformal accuracy indicates the fraction of correct predictions at the given confidence level. Conformal coverage is the fraction of substances for which valid predictions were generated at the required confidence level.	Description of the applicability domain technique	Applicability domain technique details	name	conformal prediction	Name of the applicability domain technique																																																						
description	Adding conformal features to the classifier guarantees both, a certain level of confidence and a maximum frequency of errors (misclassifications) which will be committed by the conformal predictor. The given confidence level opposes the error rate which represents the level of uncertainty that must initially be accepted and provided by the user. Conformal accuracy indicates the fraction of correct predictions at the given confidence level. Conformal coverage is the fraction of substances for which valid predictions were generated at the required confidence level.	Description of the applicability domain technique	Applicability domain technique details																																																											
name	conformal prediction	Name of the applicability domain technique																																																												
AD parameters	<table border="1"> <tr> <td>confidence</td> <td>0.8</td> <td>Conformal confidence (from 0 to 1)</td> <td>Parameters of the applicability domain technique</td> </tr> </table>	confidence	0.8	Conformal confidence (from 0 to 1)	Parameters of the applicability domain technique																																																									
confidence	0.8	Conformal confidence (from 0 to 1)	Parameters of the applicability domain technique																																																											
Goodness of fit statistics	<table border="1"> <tr> <td>Specificity</td> <td>0.83</td> <td>na</td> <td>Goodness of fit statistics</td> </tr> <tr> <td>MCC</td> <td>0.61</td> <td>na</td> <td></td> </tr> </table>	Specificity	0.83	na	Goodness of fit statistics	MCC	0.61	na																																																						
Specificity	0.83	na	Goodness of fit statistics																																																											
MCC	0.61	na																																																												
Internal validation 1	<table border="1"> <tr> <td>Conformal coverage</td> <td>0.79</td> <td>na</td> <td>Internal validation method and statistics for internal cross-validation</td> </tr> <tr> <td>Conformal accuracy</td> <td>0.81</td> <td>na</td> <td></td> </tr> </table>	Conformal coverage	0.79	na	Internal validation method and statistics for internal cross-validation	Conformal accuracy	0.81	na																																																						
Conformal coverage	0.79	na	Internal validation method and statistics for internal cross-validation																																																											
Conformal accuracy	0.81	na																																																												
Internal validation 2			Additional validation settings and statistics for internal cross-validation																																																											
External validation			External validation statistics																																																											
Comments			Additional comments on model building																																																											

Other information

[Edit](#)

	Value	Description
Other related models		Related models
Date of QMRF		Date of QMRF, usually is the same as the model creation
Date of QMRF updates		Dates of document updates
QMRF updates		Update of the QMRF
References		Reference(s) to main scientific papers and/or software package
QMRF same models		Availability of another QMRF for exactly the same model
Comment on the endpoint		Additional comments on the modeled endpoint
Endpoint data quality and variability		Text field containing all the information for adequately judge the quality of the experimental data.
Descriptor selection		Detailed information about descriptor selection process