Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Adriana de Oliveira Fernandes, Valéria Vieira Moura Paixão, Yria Jaine Andrade Santos, Eduardo Borba Alves, Ricardo Pereira Rodrigues, Daniela Aparecida Chagas-Paula, Aurélia Santos Faraoni, Rosana Casoti, Marcus Vinicius de Aragão Batista, Marcel Bermudez, Silvio Santana Dolabella and Tiago Branquinho Oliveira

Explanation regarding the models' parameters and tuning

1. Partitioning of 66%: mentioned in the manuscript.

2. Type of partitioning: The data partitioning types available in Knime are stratified sampling, draw randomly, linear sampling and take-from-top. All of them were tested for both types of algorithm (i.e. classification and regression) and for all databases and models. The partitioning type that resulted in 100% reliability in the applicability domain (APD) assessment and generated better results in terms of high accuracy or R^2 and low error indexes were thus chosen for each model.

3. Seeds: For the classification models the partitioning seed used was the number 2023, which was chosen by the authors based on the year that the project was written (i.e. 2020). It was added +1.0 to it in each round of tests and in a total of five rounds. In the end, it was noticeable that the seed 2023 generated better results in accuracy/R^2 and APD reliability than the seeds 2020, 2021, 2022 and 2024, therefore we have decided to keep this value. For the regression models the previous strategy was also used, but the results were not as adequate. Thus, we have decided to initiate the rounds with numbers from 1 to 5, being 1 the one that generated better results for FlavStr, but only for the regression decision tree (DTr). All the other regression models fared better with linear partitioning, therefore they do not require any seed.

4. Cross-validation (CV) set: For the CV set the software has available only the partitioning types stratified sampling, draw randomly, linear sampling and leave-one-out (LOO). After testing all of them in search of high accuracy or R^2 values, it was observed that the LOO type generated the best results with the default seed (i.e. 1562765658145) for all models.

5. Parameter optimisation loop: A parameter opmitisation algorithm was used in order to optimise the numeric parameters of each devised prediction algorithm. The search algorithm used was Brute Force for all procedures, while the start value was always 0 or 0.1 for parameters that needed to be higher than 0. The step size was either 0.1 or 1.0 depending if the default value was a real value or an integer, respectively. For every parameter the stop value was set for 100, with the exception of random seeds, which were set to 1000. All other boolean or non-numeric parameters were tested manually.

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S5. Parameters of the classificatory decision tree (DTc) for the FlavStr database.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for cross-validation set	1562765658145
Use binary splits only	False (default)
Do not collapse tree	False
Set confidence threshold for prunning	None
Debug (output additional information to the console)	False (default)
Minimum number of instances per leaf	9
Number of folds for reduced error prunning	None
Use reduced error prunning	False (default)
Do not clean up after the tree has been built	False (default)
Seed for random data shuffling	None
Don't perform subtree raising	True (default)
Use unprunned tree	True
Use Laplace smoothing for predicted probabilities	True
Do not use MDL correction for information gain on numeric attributes	True (default)

Table S6. Parameters of the classificatory decision tree (DTc) for the FabG model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for cross-validation set	1562765658145
Use binary splits only	False (default)
Do not collapse tree	False
Set confidence threshold for prunning	None
Debug (output additional information to the console)	False (default)
Minimum number of instances per leaf	3
Number of folds for reduced error prunning	None
Use reduced error prunning	False (default)
Do not clean up after the tree has been built	False (default)
Seed for random data shuffling	None
Don't perform subtree raising	True (default)
Use unprunned tree	True
Use Laplace smoothing for predicted probabilities	True
Do not use MDL correction for information gain on	True (default)

numeric attributes	

Table S7. Parameters of the classificatory decision tree (DTc) for the FabZ model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for cross-validation set	1562765658145
Use binary splits only	False (default)
Do not collapse tree	False
Set confidence threshold for prunning	None
Debug (output additional information to the console)	False (default)
Minimum number of instances per leaf	4
Number of folds for reduced error prunning	None
Use reduced error prunning	False (default)
Do not clean up after the tree has been built	False (default)
Seed for random data shuffling	None
Don't perform subtree raising	True (default)
Use unprunned tree	True
Use Laplace smoothing for predicted probabilities	True
Do not use MDL correction for information gain on numeric attributes	True (default)

Table S8. Parameters of the classificatory decision tree (DTc) for the FabI model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for cross-validation set	1562765658145
Use binary splits only	False (default)
Do not collapse tree	False
Set confidence threshold for prunning	None
Debug (output additional information to the console)	False (default)
Minimum number of instances per leaf	2
Number of folds for reduced error prunning	None
Use reduced error prunning	False (default)
Do not clean up after the tree has been built	False (default)
Seed for random data shuffling	None
Don't perform subtree raising	True (default)
Use unprunned tree	True

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Use Laplace smoothing for predicted probabilities	True
Do not use MDL correction for information gain on numeric attributes	True (default)

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S9. Parameters of the classificatory decision tree for the pharmacophore detection (DT_{SAR}).

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for cross-validation set	1562765658145
Use binary splits only	False (default)
Do not collapse tree	False
Set confidence threshold for prunning	None
Debug (output additional information to the console)	False (default)
Minimum number of instances per leaf	1
Number of folds for reduced error prunning	None
Use reduced error prunning	False (default)
Do not clean up after the tree has been built	False (default)
Seed for random data shuffling	None
Don't perform subtree raising	True (default)
Use unprunned tree	True
Use Laplace smoothing for predicted probabilities	True
Do not use MDL correction for information gain on numeric attributes	True (default)

Table S10. Parameters of the regression decision tree (DTr) for the FlavStr database.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Draw randomly
Partitioning random seed	1
Type of cross-validation set partitioning	Stratified sampling
Partitioning random seed for cross-validation set	1562765658145
Ignore columns without domain information	False (default)
Enable highlighting	False (default)
Use binary splits for nominal atributes	True (default)
Missing value handling	XGBoost (default)
Limit number of levels (tree depth)	4
Minimum split node size	11
Minimum node size	2

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S11. Parameters of the regression decision tree (DTr) for the FabG model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for cross-validation set	1562765658145
Ignore columns without domain information	False (default)
Enable highlighting	False (default)
Use binary splits for nominal atributes	True (default)
Missing value handling	XGBoost (default)
Limit number of levels (tree depth)	2
Minimum split node size	2
Minimum node size	1

Table S12. Parameters of the regression decision tree (DTr) for the FabZ model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for cross-validation set	1562765658145
Ignore columns without domain information	False (default)
Enable highlighting	False (default)
Use binary splits for nominal atributes	True (default)
Missing value handling	XGBoost (default)
Limit number of levels (tree depth)	10
Minimum split node size	2
Minimum node size	1

Table S13. Parameters of the regression decision tree (DTr) for the FabI model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for cross-validation set	1562765658145
Ignore columns without domain information	False (default)
Enable highlighting	False (default)
Use binary splits for nominal atributes	True (default)
Missing value handling	XGBoost (default)
Limit number of levels (tree depth)	1
Minimum split node size	7
Minimum node size	3

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S15. Parameters of the classificatory random forest (RFc) for the FlavStr database.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Split criterion	Information Gain Ratio (default)
Limit number of levels (tree depth)	1
Minimum child node size	None
Number of models	19
Use static random seed	1608215990908

Table S16. Parameters of the classificatory random forest (RFc) for the FabG model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Split criterion	Information Gain Ratio (default)
Limit number of levels (tree depth)	1
Minimum child node size	None
Number of models	13
Use static random seed	1608215990908

Table S17. Parameters of the classificatory random forest (RFc) for the FabZ model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Split criterion	Information Gain Ratio (default)
Limit number of levels (tree depth)	1
Minimum child node size	None
Number of models	2
Use static random seed	1608215990908

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S18. Parameters of the classificatory random forest (RFc) for the FabI model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Split criterion	Information Gain Ratio (default)
Limit number of levels (tree depth)	10
Minimum child node size	1
Number of models	4
Use static random seed	1608215990908

Table S19. Parameters of the regression random forest (RFr) for the FlavStr database.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Draw randomly
Partitioning random seed	1
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Limit number of models (tree depth)	6
Minimum child node size	1
Number of models	29
Use static random seed	1608215990908

Table S20. Parameters of the regression random forest (RFr) for the FabG model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Limit number of models (tree depth)	10
Minimum child node size	None
Number of models	26
Use static random seed	1608215990908

Table S21. Parameters of the regression random forest (RFr) for the FabZ model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Limit number of models (tree depth)	10
Minimum child node size	None
Number of models	35

Use static random seed	1608215990908

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S22. Parameters of the regression random forest (RFr) for the FabI model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Limit number of models (tree depth)	10
Minimum child node size	None
Number of models	11
Use static random seed	1608215990908

Table S23. Parameters of the classificatory support vector machine (SVMc) for the FlavStr database.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Type of SVM	C-SVC (default)
Type of kernel	Radial basis function (default)
Degree in kernel function	3 (default)
Set gamma in kernel function	1 (default)
Set coeficient value	0 (default)
Cost	1.0 (default)
Set nu value	0.5 (default)
Normalise input data	True
Nominal do binary conversion	True (default)
Missing value replacement	True (default)
Set the epsilon in loss function	None
Shrinking heuristics	True (default)
Generate probability estimates for classification	True
Random seed	1 (default)

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S24. Parameters of the classificatory support vector machine (SVMc) for the FabG model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Type of SVM	Nu-SVC
Type of kernel	Radial basis function (default)
Degree in kernel function	3 (default)
Set gamma in kernel function	14
Set coeficient value	0 (default)
Cost	1.0 (default)
Set nu value	0.5 (default)
Normalise input data	True
Nominal do binary conversion	True (default)
Missing value replacement	True (default)
Set the epsilon in loss function	None
Shrinking heuristics	True (default)
Generate probability estimates for classification	True
Random seed	1 (default)

Table S25. Parameters of the classificatory support vector machine (SVMc) for the FabZ model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Type of SVM	C-SVC (default)
Type of kernel	Polynomial
Degree in kernel function	4
Set gamma in kernel function	1 (default)
Set coeficient value	0 (default)
Cost	1.0 (default)
Set nu value	None
Normalise input data	True
Nominal do binary conversion	True (default)
Missing value replacement	True (default)
Set the epsilon in loss function	None
Shrinking heuristics	True (default)
Generate probability estimates for classification	True
Random seed	1 (default)

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S26. Parameters of the classificatory support vector machine (SVMc) for the FabI model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Type of SVM	C-SVC (default)
Type of kernel	Radial basis function (default)
Degree in kernel function	2
Set gamma in kernel function	1 (default)
Set coeficient value	2
Cost	1 (default)
Set nu value	None
Normalise input data	True
Nominal do binary conversion	True (default)
Missing value replacement	True (default)
Set the epsilon in loss function	None
Shrinking heuristics	True (default)
Generate probability estimates for classification	True
Random seed	1 (default)

Table S27. Parameters of the regression support vector machine (SVMr) for the FlavStr database.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Draw randomly
Partitioning random seed	1
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Type of SVM	Nu-SVR
Type of kernel	Linear
Degree in kernel function	3 (default)
Set gamma in kernel function	0
Set coeficient value	0 (default)
Cost	1 (default)
Set nu value	0.46
Normalise input data	True
Nominal do binary conversion	True (default)
Missing value replacement	True (default)
Set the epsilon in loss function	0.1 (default)
Shrinking heuristics	True (default)
Generate probability estimates for classification	True
Random seed	1 (default)

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S28. Parameters of the regression support vector machine (SVMr) for the FabG model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Random sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Type of SVM	Epsilon-SVR
Type of kernel	Linear
Degree in kernel function	3 (default)
Set gamma in kernel function	1 (default)
Set coeficient value	0 (default)
Cost	10
Set nu value	None
Normalise input data	True
Nominal do binary conversion	True (default)
Missing value replacement	True (default)
Set the epsilon in loss function	0.1 (default)
Shrinking heuristics	True (default)
Generate probability estimates for classification	True
Random seed	1 (default)

Table S29. Parameters of the regression support vector machine (SVMr) for the FabZ model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Random sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Type of SVM	Nu-SVR
Type of kernel	Linear
Degree in kernel function	3 (default)
Set gamma in kernel function	1 (default)
Set coeficient value	0 (default)
Cost	16
Set nu value	0.44
Normalise input data	True
Nominal do binary conversion	True (default)
Missing value replacement	True (default)
Set the epsilon in loss function	None
Shrinking heuristics	True (default)
Generate probability estimates for classification	True
Random seed	1 (default)

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S30. Parameters of the regression support vector machine (SVMr) for the FabI model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Random sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Type of SVM	Nu-SVR
Type of kernel	Linear
Degree in kernel function	3 (default)
Set gamma in kernel function	1 (default)
Set coeficient value	0 (default)
Cost	8
Set nu value	0.66
Normalise input data	True
Nominal do binary conversion	True (default)
Missing value replacement	True (default)
Set the epsilon in loss function	None
Shrinking heuristics	True (default)
Generate probability estimates for classification	True
Random seed	1 (default)

Table S31. Parameters of the classificatory multilayer perceptron (MLPc) for the FlavStr database.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Auto build	True (default)
Debug	False (default)
Decay	False (default)
Hidden layers	2 (4 neurons each)
Learning rate for the backpropagation	0.3 (default)
Momentum rate for the backpropagation	0.1
Nominal to binary filter	True (default)
Normalize attributes	True (default)
Normalize numeric class	True (default)
Reset	False (default)
Seed	2023
Number of epochs	400
Validation set size to terminate training (%)	0 (default)
Consecutive number of errors allowed for validation testing before network terminates	20 (default)

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S32. Parameters of the classificatory multilayer perceptron (MLPc) for the FabG model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Auto build	True
Debug	False
Decay	False
Hidden layers	2 (4 neurons each)
Learning rate for the backpropagation	0.32
Momentum rate for the backpropagation	0.25
Nominal to binary filter	True
Normalize attributes	True
Normalize numeric class	True
Reset	False
Seed	6
Number of epochs	430
Validation set size to terminate training (%)	0 (default)
Consecutive number of errors allowed for validation testing before network terminates	20 (default)

Table S33. Parameters of the classificatory multilayer perceptron (MLPc) for the FabZ model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Auto build	True
Debug	False
Decay	False
Hidden layers	2 (4 neurons each)
Learning rate for the backpropagation	0.19
Momentum rate for the backpropagation	0.21
Nominal to binary filter	True
Normalize attributes	True
Normalize numeric class	True
Reset	False
Seed	0
Number of epochs	440
Validation set size to terminate training (%)	0 (default)
Consecutive number of errors allowed for validation testing before network terminates	20 (default)

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S34. Parameters of the classificatory multilayer perceptron (MLPc) for the FabI model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Auto build	True
Debug	False
Decay	False
Hidden layers	2 (4 neurons each)
Learning rate for the backpropagation	0.2
Momentum rate for the backpropagation	0.22
Nominal to binary filter	True
Normalize attributes	True
Normalize numeric class	True
Reset	False
Seed	33
Number of epochs	443
Validation set size to terminate training (%)	0 (default)
Consecutive number of errors allowed for validation testing before network terminates	20 (default)

Table S35. Parameters of the regression multilayer perceptron (MLPr) for the FlavStr database.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Draw randomly
Partitioning random seed	1
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Auto build	True (default)
Debug	False (default)
Decay	False (default)
Hidden layers	2 (4 neurons each)
Learning rate for the backpropagation	0.05
Momentum rate for the backpropagation	0.18
Nominal to binary filter	True (default)
Normalize attributes	True (default)
Normalize numeric class	True (default)
Reset	False (default)
Seed	461
Number of epochs	276
Validation set size to terminate training (%)	0 (default)
Consecutive number of errors allowed for validation testing before network terminates	20 (default)

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S36. Parameters of the regression multilayer perceptron (MLPr) for the FabG model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Auto build	True
Debug	False
Decay	False
Hidden layers	2 (4 neurons each)
Learning rate for the backpropagation	0.15
Momentum rate for the backpropagation	0.22
Nominal to binary filter	True
Normalize attributes	True
Normalize numeric class	True
Reset	False
Seed	534
Number of epochs	286
Validation set size to terminate training (%)	0 (default)
Consecutive number of errors allowed for validation testing before network terminates	20 (default)

Table S37. Parameters of the regression multilayer perceptron (MLPr) for the FabZ model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Auto build	True
Debug	False
Decay	False
Hidden layers	2 (4 neurons each)
Learning rate for the backpropagation	0.2
Momentum rate for the backpropagation	0.2 (default)
Nominal to binary filter	True
Normalize attributes	True
Normalize numeric class	True
Reset	False
Seed	602
Number of epochs	275
Validation set size to terminate training (%)	0 (default)
Consecutive number of errors allowed for validation testing before network terminates	20 (default)

Table S38. Parameters of the regression multilayer perceptron (MLPr) for the FabI model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Auto build	True
Debug	False
Decay	False
Hidden layers	2 (4 neurons each)
Learning rate for the backpropagation	0.05
Momentum rate for the backpropagation	0.79
Nominal to binary filter	True
Normalize attributes	True
Normalize numeric class	True
Reset	False
Seed	35
Number of epochs	356
Validation set size to terminate training (%)	0 (default)
Consecutive number of errors allowed for validation testing before network terminates	20 (default)

Table S39. Parameters of the partial least squares (PLS) for the FlavStr database.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Draw randomly
Partitioning random seed	1
Type of cross-validation set partitioning	Leave-one-out
Train control method	None
Number of components	4

Table S40. Parameters of the partial least squares (PLS) for the FabG model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear smapling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Train control method	None
Number of components	3

Table S41. Parameters of the partial least squares (PLS) for the FabZ model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Train control method	None

Number of components	3

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S42. Parameters of the partial least squares (PLS) for the FabI model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Train control method	None
Number of components	3

Table S43. Parameters of the k-nearest neighbour (KNN) for the FlavStr database.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Stratified sampling
Partitioning random seed	2023
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Number of nearest neighbours	13
Cross validate	False (default)
Debug	False (default)
Distance weighting method	Weight by 1/distance (default)
Use mean squared error rather than mean absolute error when doing cross-validation for regression problems	False (default)
Nearest neighbour search algorithm	LinearNNSearch (default)
Window size	50

Table S44. Parameters of the k-nearest neighbour (KNN) for the FabG model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Number of nearest neighbours	5
Cross validate	False (default)
Debug	False (default)
Distance weighting method	Weight by 1/distance (default)
Use mean squared error rather than mean absolute error when doing cross-validation for regression problems	False (default)
Nearest neighbour search algorithm	LinearNNSearch (default)
Window size	6

Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - parameters for the construction of the models and their tuning

Table S45. Parameters of the k-nearest neighbour (KNN) for the FabZ model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Number of nearest neighbours	4
Cross validate	False (default)
Debug	False (default)
Distance weighting method	Weight by 1/distance (default)
Use mean squared error rather than mean absolute error when doing cross-validation for regression problems	False (default)
Nearest neighbour search algorithm	LinearNNSearch (default)
Window size	6

Table S46. Parameters of the k-nearest neighbour (KNN) for the FabI model.

Parameters	Values
Training set partitioning	66%
Type of training set partitioning	Linear sampling
Partitioning random seed	None
Type of cross-validation set partitioning	Leave-one-out
Partitioning random seed for the cross-validation set	1562765658145
Number of nearest neighbours	7
Cross validate	False (default)
Debug	False (default)
Distance weighting method	Weight by 1/distance (default)
Use mean squared error rather than mean absolute error when doing cross-validation for regression problems	False (default)
Nearest neighbour search algorithm	LinearNNSearch (default)
Window size	7