CIRCE: Web-Based Platform for the Prediction of Cannabinoid Receptor Ligands Using Explainable Machine Learning

Nicola Gambacorta[1,2], Fulvio Ciriaco[3], Nicola Amoroso[1], Cosimo Damiano Altomare[1], Jürgen Bajorath[2] and Orazio Nicolotti[1]**

[1]Dipartimento di Farmacia Scienze del Farmaco, Università degli Studi di Bari "Aldo Moro", Via E. Orabona, 4, I-70125 Bari, Italy;

[2]Department of Life Science Informatics and Data Science, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Friedrich-Hirzebruch-Allee 5/6, D-53115 Bonn, Germany;

[3]Dipartimento di Chimica, Università degli Studi di Bari "Aldo Moro", Via E. Orabona, 4, I-70125 Bari, Italy;

*Corresponding authors

orazio.nicolotti@uniba.it (for editorial correspondence)

bajorath@bit.uni-bonn.de

ABSTRACT

The endocannabinoid system, which includes cannabinoid receptor 1 and 2 subtypes (CB_1R and CB_2R, respectively), is responsible for the onset of various pathologies including neurodegeneration, cancer, neuropathic and inflammatory pain, obesity, and inflammatory bowel disease. Given the high similarity of CB_1R and CB_2R, generating subtype-selective ligands is still an open challenge. In this work, the Cannabinoid Iterative Revaluation for Classification and Explanation (CIRCE) compound prediction platform has been generated based on explainable machine learning to support the design of selective CB_1R and CB_2R ligands. Multi-layer classifiers were combined with Shapley value analysis to facilitate explainable predictions. In test calculations, CIRCE predictions reached ~80% accuracy and structural features determining ligand predictions were rationalized. CIRCE was designed as a web-based prediction platform that is made freely available as a part of our study.

Introduction

Cannabinoid receptors 1 and 2 (CB_1R and CB_2R) constitute the endocannabinoid system and represent the molecular targets of the 9-tetrahydrocannabinol (9-THC), a psychoactive agent derived from *Cannabis sativa*. CB_1R and CB_2R are responsible for many physiological functions such as appetite, pain perception, memory, and immunomodulation.[1,2]

CB_1R and CB_2R are largely expressed in the central nervous system (CNS) as well as in the immune system and have distinct tissue distributions and functions. CB_1R is a major player in the regulation of higher cognitive functions, neuronal development and synaptic plasticity, reward and addiction, pain, and food intake. CB_1R is also associated with biological and pathological processes outside the CNS, being its expression reported in different types of hepatic cells, in the cardiovascular system, in the adipose tissue, muscles, and mitochondria. The CB_1R deregulation is behind the onset of several pathological conditions such as obesity[3–5], neurodegenerative diseases, glaucoma, pain, and cancer. Unlike CB_1R, CB_2R has so far received less attention, and when it was first discovered, CB_2 activity was only found in lymphoid organs, immune cells, and hematopoietic cells. In fact, CB_2 is primarily expressed in all immune system tissues and circulating cells, with varying degrees of expression and activity depending on the stimulus, cell type, and cell activation. In this respect, CB_2R plays a pivotal role in a wide spectrum of pathological conditions: it can act as an antitumor agent by inhibiting cells proliferation or by decreasing angiogenesis or metastasis, or it can be used for palliative care [6]; furthermore, it is also implicated in several central nervous system conditions.[1,7–9]

CB_1R and CB_2R are closely related subtypes and share ~68% sequence homology in the transmembrane region and ~44% overall. Accordingly, the generation of subtype-selective ligands is extremely difficult. To the best of our knowledge, the ~100 most popular synthetic cannabinoids that act as CB_1R and CB_2R agonists fall into different chemical categories: classical, nonclassical, amino-alkylindole, eicosanoids, and others. [10–12] The classical family of CB_1R and CB_2R agonists is constituted by the dibenzopyran derivatives. Two especially notable examples are (Δ)-9-

tetrahydrocannabinol (Δ9-THC), the primary psychoactive component of cannabis, and (Δ)-11-hydroxy-8-THC-dimethylheptyl (HU-210) [13], a synthetic analog of (Δ)-8-THC. On the other hand, the nonclassical category includes bicyclic and tricyclic analogs of Δ9-THC devoid of the pyran ring, with CP-55,940 being one of its well-known members.[14] The compounds belonging to the amino-alkylindole class (i.e. WIN-55,212-2) considerably differ from the classical and non-classical cannabinoid receptor ligands. [14,15] For completeness, the structures of the mentioned compounds are reported in Figure 1.

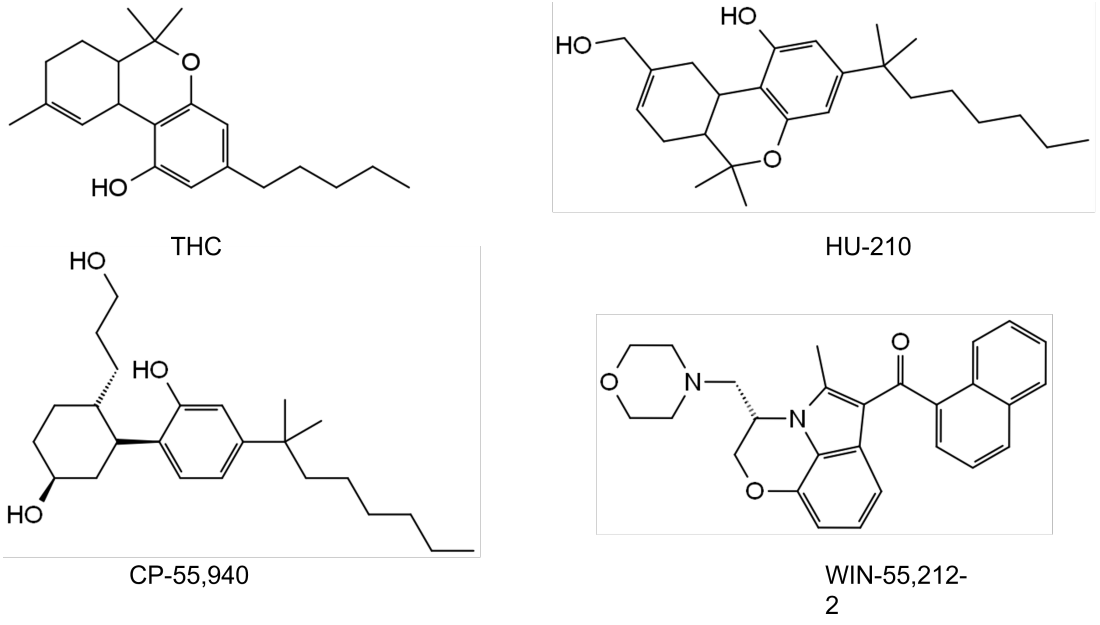


Figure 1. Examples of known cannabinoid receptors ligands.

We have set out to derive multi-layer explainable machine learning (XML) models[16–19] for predicting CB_1R and CB_2R ligands. Specifically, four cooperating classification models were generated for predicting and rationalizing molecular determinants driving selective ligand binding to CB_1R and CB_2R. Each classification model was independently derived based on different sets of training data carefully curated from the ChEMBL database (release 31). [20] These compound pools contain a wealth of chemical information along with high-quality experimental data for binding to

CB_1R and CB_2R. A new substructure-based core-substituent fingerprint (CSFP)[21] was used to encode the structural information and SHAP values[22–24] were computed to explain individual predictions. Eventually, the four models were assembled to build a multi-layer classifier[25] which is freely accessible through a web platform designated Cannabinoid Iterative Revaluation for Classification and Explanation (CIRCE) that is provided with a user-friendly graphical interface.[26,27] CIRCE returns predictions on demand and instantly provides a detailed portable report of prediction outcomes. While a variety of studies have reported compound predictions for the cannabinoid receptor system[28–38], to the best of our knowledge, CIRCE is the first free web platform enabling users to predict if a query compound might interact with CB_1R or CB_2R. Moreover, CIRCE's XML framework provides model predictions and easy to understand color-coded maps of feature mapping to test compounds.

Results and Discussion

The multi-layer RF model forming the core component of CIRCE was conceived for the identification of selective CB_1R or CB_2R ligands by applying increasingly stringent criteria to discriminate between potential ligands and other compounds. Therefore, four individual models were derived to act sequentially by addressing subsequent prediction tasks for query compounds. The predictions were then explained via SHAP analysis to identify features determining the prediction and study structural motifs in selective ligands formed by decisive features. Notably, CIRCE was designed to predict if unknown compounds can act as CB_1R or CB_2R ligands but does not distinguish between agonists and antagonists. This is the case because many candidate ligands with potency measurements had no or not clearly defined mode-of-action annotations, which prohibited meaningful mechanism-based model derivation. Thus, we preferred instead to employ this amount of available data as an external set to strengthen the generalization of the multi-layer classifiers in CIRCE.

As a first step, Analog Series (AS) were identified in the selective (including the D_{CB1} and D_{CB2} collections of 1477 CB_1R and 1820 CB_2R specific ligands, respectively) and non-selective (including the D_{MT} collection of 1251 non-selective CB_1R and CB_2R ligands) datasets. The Venn diagram [39] in Figure 2 reports the number of AS identified in these datasets and their overlap. Each AS contains a unique core structure.

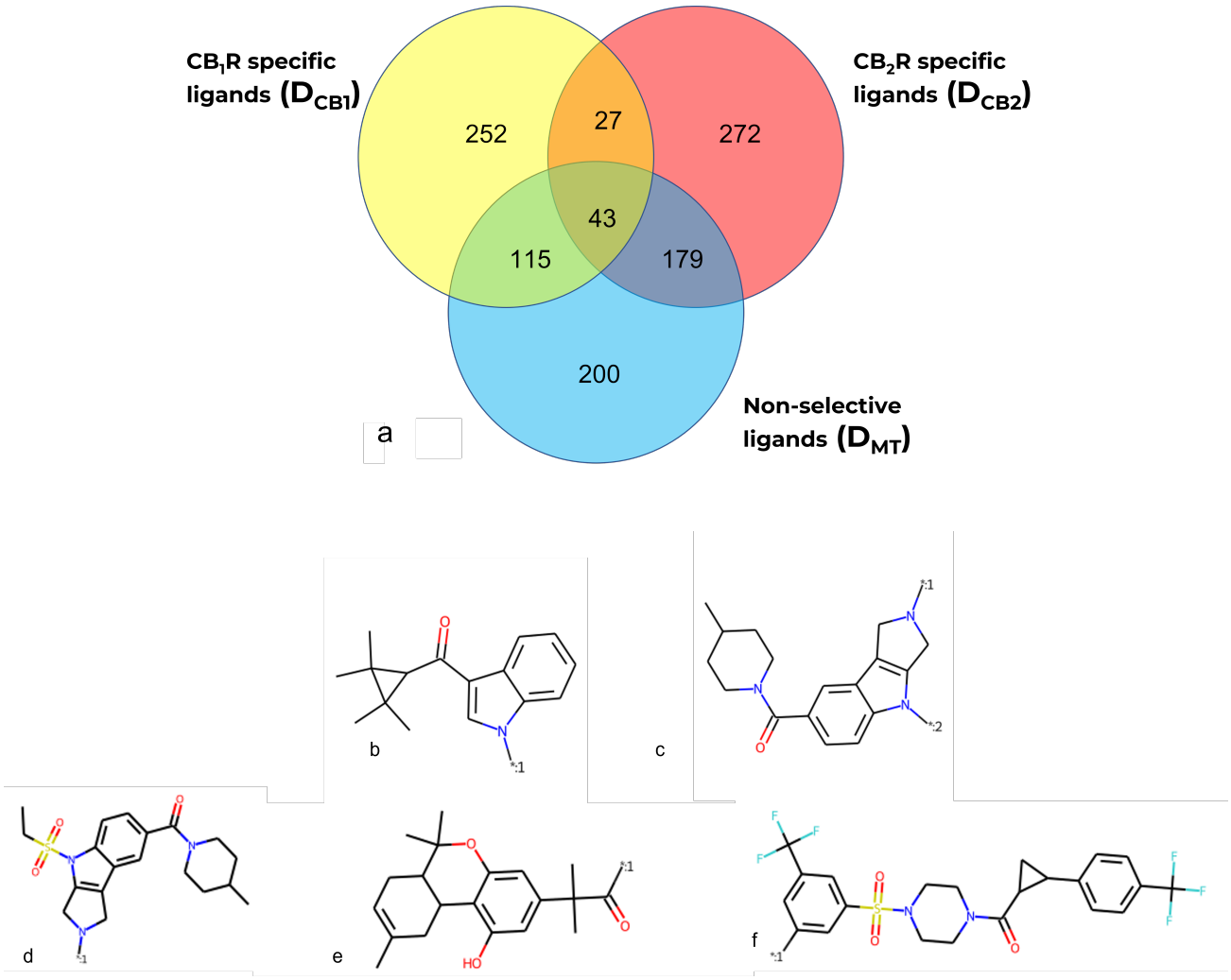


Figure 2. Panel (a) The Venn diagram shows the extent of chemical core overlap for selective and unselective CB_1R and CB_2R ligands. The yellow and red circles indicate selective ligands while the cyan circle indicates non-selective ligands. Panels from (b) to (f) show representative AS cores

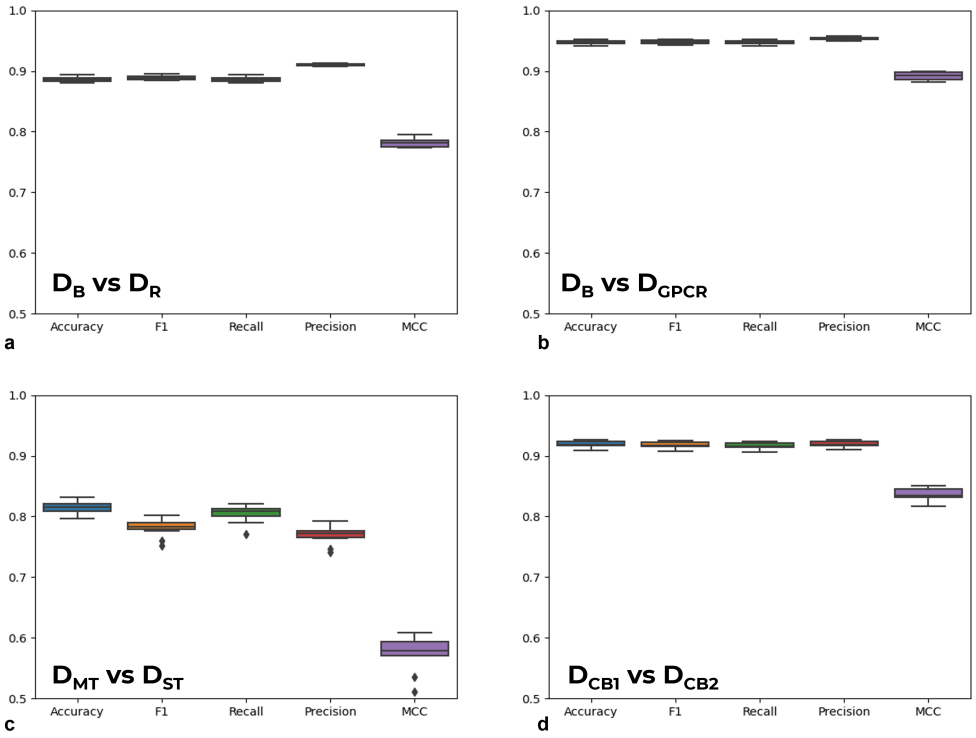
taken from selective and non-selective datasets. Symbols "*:1" and "*:2" indicate the position where the substituents occur.

We identified 252, 272 and 200 cores that only occurred in the D_{CB1}, D_{CB2} and D_{MT} datasets, respectively. The intersection $D_{CB1} \cap D_{CB2}$ revealed the presence of cores important for binding to both CB_1R and CB_2R, indicating non-selectivity. The intersections $D_{CB1} \cap D_{MT}$, and $D_{CB2} \cap D_{MT}$ contained cores with preferentially binding to CB_1R or CB_2R, respectively. On the other hand, cores falling in the intersection $D_{CB1} \cap D_{CB2} \cap D_{MT}$ should be considered unselective. Differentiating between cores in selective and non-selective cannabinoid ligands is relevant for drug design. For instance, core 1 (panel (b) of Figure 2), composed of the 1H-indol-3-yl-(2,2,3,3-tetramethylcyclopropyl) methanone moiety, was found in D_{CB1}, D_{CB2}, and D_{MT} datasets 20, 6 and 1 times, respectively, hence representing an unselective core. Furthermore, cores 2 and 3 (panels (c) and (d) of Figure 2), shared the 2-methyl-1,2,3,4-tetrahydropyrrolo[3,4-B]indole and the acetylpiperidine moieties, and differed only for the absence/presence of the ethylsulphone group. Interestingly, this small modification depicted for cores 2 and 3 was implicated in selectivity, since their occurrences were retrieved within D_{CB2} and D_{MT} datasets, respectively, and only within D_{CB1} dataset, respectively. Moreover, cores 4 and 5, shown in panels (e), and (f) of Figure 2, were found only in the D_{CB2}, and D_{MT} datasets, respectively.

Model performance

We next evaluated the Random Forest (RF) models on the basis of different performance measures in independent trials. The performance evaluation of the four independent RF models based on a 10-fold cross validation is summarized in Figure 3.

Based on a sampling in the range 0.1 to 0.9 with a step equal to 0.1 (as shown in Figure S1 of the Supporting Information), the cut-off values of the classification scores were set to 0.4, 0.4, 0.6 and 0.5, for the first, second, third and fourth layer, respectively, to maximize the Matthews Correlation Coefficient (MCC) yield.



136

137 **Figure 3.** The panels (a), (b), (c) and (d) show the performances of the four independent classifiers
138 after a 10-fold cross validation based on: Ligand collection (D_B) *vs* Random Collection (D_R);
139 Ligand collection (D_B) *vs* GPCR collection (D_{GPCRs}); Non-selective ligand collection (D_{MT}) *vs*
140 Selective ligand collection (D_{ST}); and CB_1R collection (D_{CB1}) *vs* CB_2R collection (D_{CB2}). Accuracy,
141 F1, recall, precision and MCC values are used as metrics.

142 Overall, the predictions were accurate and stable with very small differences over the 10
143 independent trials, as clearly described by the narrow distributions depicted in the box plots of
144 Figure 3. Prediction accuracy was consistently beyond 85% for the first two models and the fourth
145 model. Furthermore, selective and non-selective CB_1R/CB_2R ligands were distinguished with
146 greater than 80% accuracy by the third model. Notably, the final model differentiated between
147 selective CB_1R from selective CB_2R ligands with greater than 90% accuracy. For the sake of
148 completeness, these data were also provided in the Table S1 of the Supporting Information.
149 Although our datasets included compounds with only K_i and IC_{50} experimental data, we challenged
150 the generalization capability of the multi-layer model by employing as an external set an unrelated

pool of 1860 cannabinoid receptor compounds with available EC_{50} values only for CB_1R or CB_2R (342 and 1518, respectively). As a result, a subset of 444 ligands (144 CB_1R and 300 CB_2R ligands) passed through the four predictions and were satisfactorily predicted to be selective CB_1R or CB_2R ligands with 78% accuracy and an MCC value of 0.55.

On the other hand, a possible reason for the relatively small number of ligands passing the multi-layer model might be their overall structural diversity with respect to the D_B dataset, as assessed by computing maximum Tanimoto similarity (Figure S2 of the Supporting Information). For the sake of completeness, we repeated the generalization study on the external set by arbitrarily lowering the cut-off values for the first two models to less stringent thresholds equal to 0.2 and 0.3. As shown in Table S1 of the Supporting Information, such adaptations of the cut-off values from 0.4 to 0.2 and from 0.4 to 0.3 increased the number of passing ligands from 444 of 1860 to 1214 and 726, respectively. This option was made available with the intention of giving users a broader view of the predictions although those made for the additional ligands at lower cut-off values should be considered with caution as further discussed below.

Model explanation and features mapping

Fingerprint features contributing to individual predictions were ranked on the basis of calculated SHAP values, described in the Materials and Methods section. Since the features used for model derivation were unique structural fragments, they were readily interpretable with respect to CB_1R or CB_2R ligand selectivity. Complementing this wealth of information with SHAP analysis, CIRCE was able to return an intuitive and explainable knowledge basis for interpreting the molecular determinants behind the selectivity towards CB_1R or CB_2R thus providing transparent and immediate clues for designing new promising ligands.

Some representative examples of features driving the prediction of CB_1R or CB_2R ligands are shown in Figure 4.

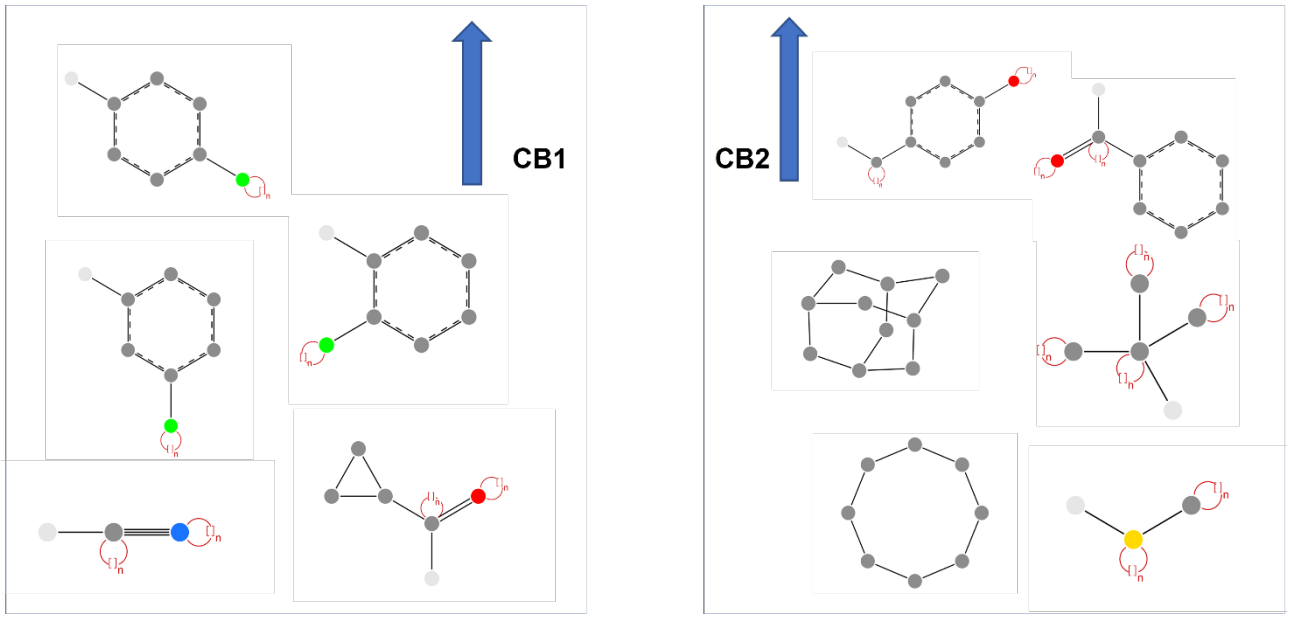


Figure 4. Representative examples of relevant substructures prioritized by SHAP analysis. Structural fragments determining the correct prediction of selective CB_1R and CB_2R ligands are shown in the boxes on the left and right, respectively. Gray, blue, green, red and yellow circles indicate carbon, nitrogen, chlorine, oxygen and sulfur atoms, respectively.

Intriguingly, the occurrence of substituents such as the acetylcyclopropanil and the cyan group as well as of phenyl rings with chlorine substituents in *para, meta,* or *ortho* positions, the latter present in well-known CB_1R selective ligands Ibipinabant and Rimonabant (a withdrawn drug previously used as anorectic antiobesity agent) [40–42], indicated selectivity for CB_1R. On the other hand, rings such as the cyclooctane and the adamantane, the latter contained in the selective CB_2R antagonist AM-10257, depicted also as cognate ligand within the CB_2 receptor crystal structure [43], and a substituent such as the *ter*-butyl were relevant for CB_2R selectivity.

This analysis was automatically included as a final step of the workflow, enabling an intuitive graphical explanation of pivotal features extracted through the SHAP analysis. Some representative examples of correctly predicted compound heat maps are shown in Figure 5. Figure 5a and 5b show correctly predicted CB_1R ligands, and substructures driving the prediction were highlighted with a gradient-based orange color employing the SHAP values as discussed above. Similarly, Figure 5c and 5d showed

correctly predicted CB_2R ligands, and the blue gradient-based color highlights the most important features for the right prediction. It is important to point out that the darker the color, the more important the substructure for the prediction.

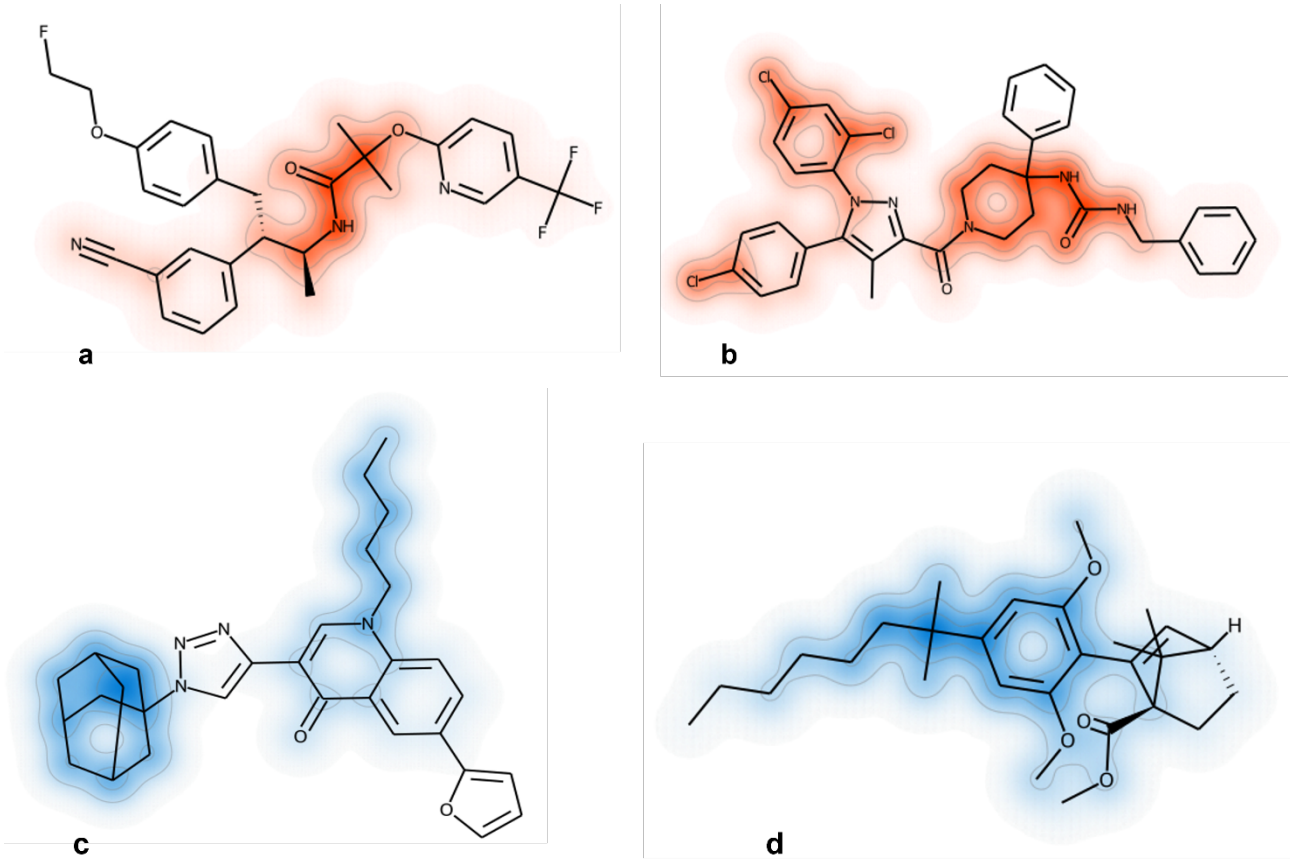


Figure 5. The panels (a) and (b) report two examples of correctly predicted CB_1R ligands. Conversely, panels (c) and (d) show two properly predicted CB_2R ligands. Orange and blue colors highlight important substructures for the prediction of CB_1R or CB_2R ligands, respectively.

CIRCE web platform

The multi-layer ML classifier and SHAP analysis were implemented in a user-friendly free web platform to provide a transparent and affordable tool for both expert and nonexpert researchers available at http://prometheus.farmacia.uniba.it/circe/.

On the "Prediction" page, users can interrogate the CIRCE platform by drawing the 2D structure of a query molecule, or by pasting SMILES code. MOL and SDF formats are also supported.

Computations take a few seconds to return an HTML output with all the required information. All the steps for CIRCE "Prediction" request are summarized in Figure 6.

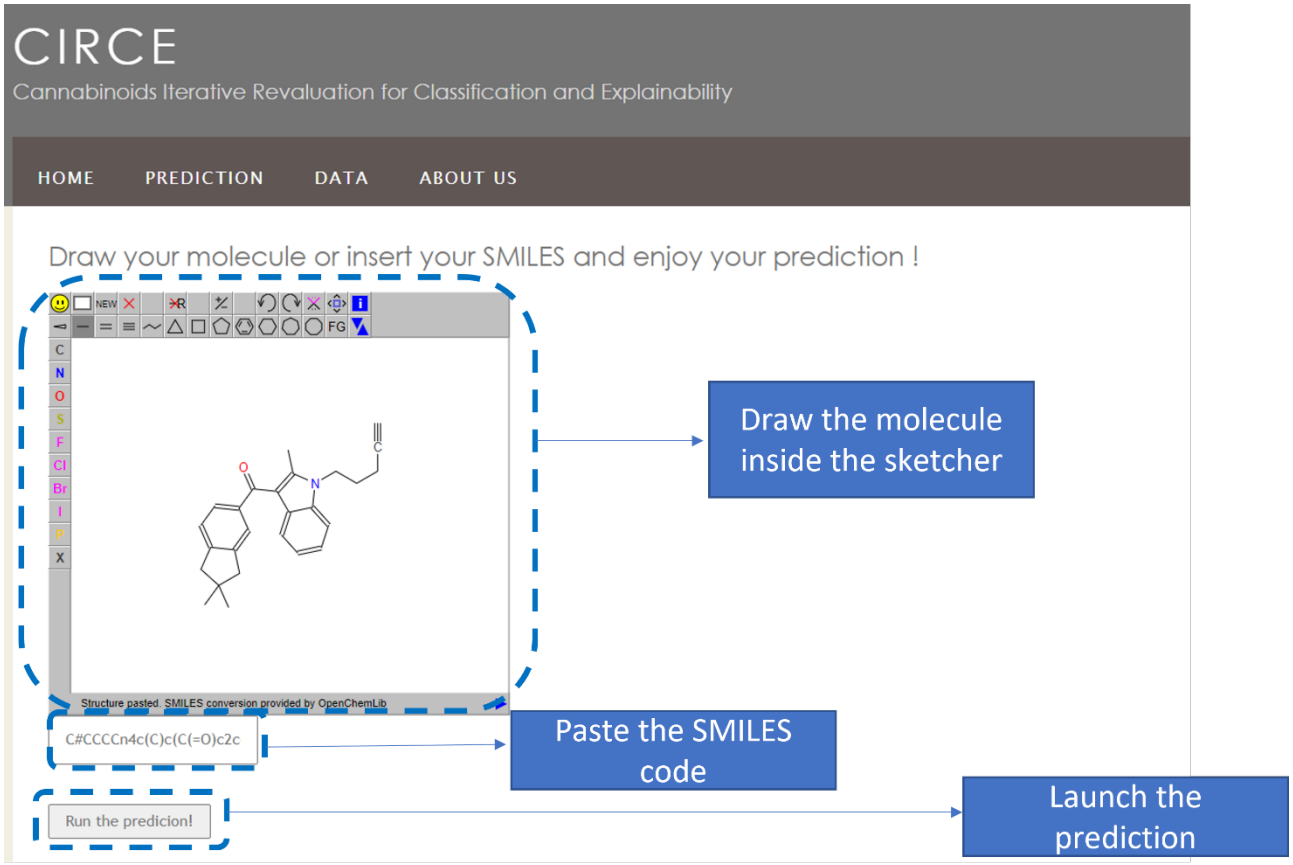


Figure 6. CIRCE "Prediction" page: query molecule can be entered as SMILES string or drawn within the JMSE sketcher. "Run the prediction!" button is to launch the prediction.

As far as the HTML output is concerned, users can retrieve full details concerning the prediction for a given query structure. In detail, the output is organized as shown in Figure 7, showing a case study fully discussed in the next section.

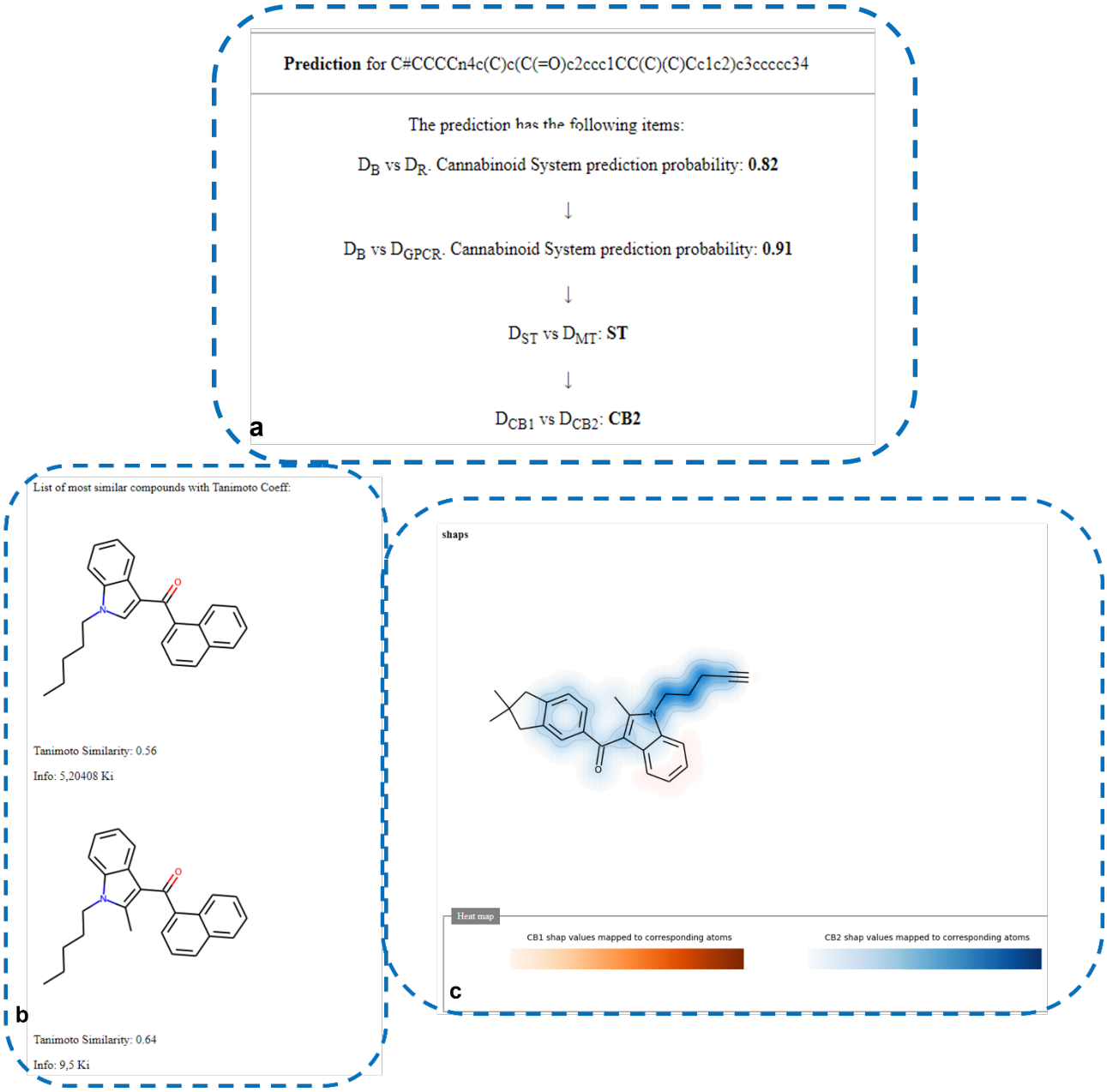


Figure 7. CIRCE output. Panels (a), (b), and (c) report the output of each machine learning model, the list of similar structures related to the query compound, and the heat-map related to the SHAP value computed on the query, respectively.

The first section gives information concerning the prediction output. For models of the first and second layers, the probability prediction of the "Cannabinoid System" class is expressed with a value ranging from 0 to 1. This value measures the likeness of a query compound to be predicted as cannabinoid ligand. More in details, if the prediction for the first layer is calculated between 0.0 and

0.2, the computation immediately stops since the query will not be labeled as a cannabinoid system ligand. If the prediction for the first two layers lies between 0.2 and 0.4, the query will be forwarded to the next layer but flagged as poorly reliable. This precautionary threshold range was selected according to the MCC values shown in panels (a) and (b) of Figure S1 and implemented to smooth the rejection of still unexplored chemotypes; this adaptation proved useful when screening external compounds as it allowed the prediction of even suspicious queries leaving the user the option to make more informed assessments. The model of the third layer reports prediction for selective (ST) or non-selective (MT) ligands. Finally, the model of the fourth layer returns assignment for selective CB_1R or CB_2R ligands.

The second section returns all similar structures belonging to the Cannabinoid System dataset compared to the query molecules, along with their experimental activity and the Tanimoto similarity coefficient.

The last section summarized the SHAP analysis for the query molecule and provides the heatmap visualization. Orange and blue colors indicate important substructures for the prediction of CB_1R and CB_2R, respectively. The darker the color, the higher the substructure importance for the prediction.

Case study

The CIRCE web platform was evaluated with literature examples of ligands not included in model derivation. For example, a selective CB_2R ligand **49**, which showed a Selectivity Index equal to 30.5. [44] As reported in panel (a) of Figure 8, the pentynyl chain along with the benzoyl groups are fundamental for the correct prediction.

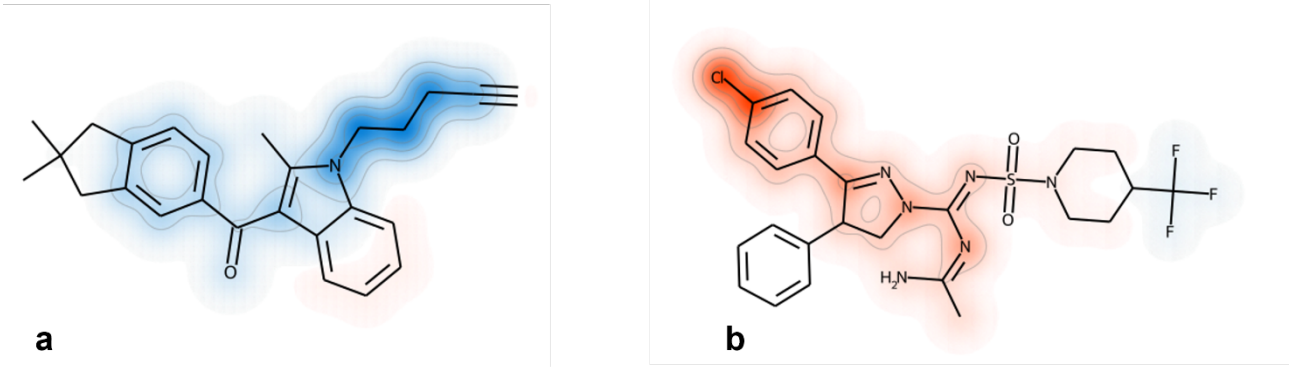


Figure 8. Panels (a) and (b) reports the structures of ligands **49** and **7f** along with the heat-map indicating the key determinants for the correct prediction.

In 2022, Iyer *et al.* tested a series of Ibipinabant based compounds for targeting CB_1R.[45] Among all molecules, compound **7c** showed a Ki value 14 nM against CB_2R and >100 nM against CB_2R. CIRCE correctly predicted these compounds and highlighted pivotal substructures for the predictions, as shown in panel (b) of Figure 8. Here, the *para*-chlorine ring and the 3,4-diarylpyrazoline ring connected to the carboximidamide group contributed the most to the CB_1R correct prediction.

Data and software availability

The CIRCE platform makes automated predictions based as a whole on 24548 small molecules provided with experimental bioactivity data concerning the endocannabinoid system. Data were selected from ChEMBLdb release 31 by employing the set of filtering rules described in the Materials and Methods section. CIRCE is written in Python and is crafted as a freely available web platform at http://prometheus.farmacia.uniba.it/circe/. All the data are available as Supporting Information.

Conclusions

The endocannabinoid system represents a major pharmaceutical target. However, given the similarity of cannabinoid receptor isoforms, the rational design of selective CB_1R and CB_2R ligands is a particularly challenging task. In this respect, to aid in the discovery of selective CB_1R and

CB_2R ligands, we have generated a multi-layer ML model to select candidate compounds with increasing stringency and complemented the predictions with SHAP analysis for model explanation and the identification of characteristic substructures in isoform-selective ligands. In test calculations, overall accurate predictions were obtained, and SHAP analysis consistently identified structural fragments determining the predictions. The XML system is provided as a freely available wed-based prediction and analysis platform. To the best of our knowledge, although several *in silico* approaches have been so far developed for the discovery of potential cannabinoid ligands, CIRCE is the first freely available digital platform enabling the transparent prediction of the selectivity against cannabinoid isoforms by employing a recently published fragment-based fingerprint. The user is given the chance to easily identifying the crucial molecular determinants involved in the classification process through an intuitive heat color map. All these steps are automatically included in the platform workflow, thus allowing broader employment for both experts and non-experts.

We hope that CIRCE will be useful to support the generation of selective CB_1R and CB_2R ligands in the practice of medicinal chemistry. In addition, CIRCE can be easily adapted on demand to run even massive virtual screening campaigns of large commercial library of chemical and natural compounds by providing a computer readable output easily transferable to the most common statistics and molecular tool for further and more informed analysis.

Experimental section

Datasets for model building

Compounds and activity data were extracted from release 31 of the ChEMBL database using the following filters [28]: (i) target filter ("target type: Single protein | Protein complex"); (ii) ligand filter ("molecule type: Small molecule"; "prodrug: not 1"); and (iii) activity record filters ("confidence score: >5"; "standard relation: ="; "standard type: Ki and IC_{50}"; "standard units: nM"; “no comment inherent to inactivity”). As reported elsewhere, [26,27,46,47] these filtering criteria ensuring high activity

295 data integrity were successfully employed for the construction of a computational tool for drug
296 target fishing and bioactivity prediction. In the present work, only data for CB_1R and CB_2R binding
297 were taken into consideration. Hence, the classifiers derived from these data predict active
298 compounds but do not differentiate between agonistic and antagonistic modes-of-action.
299 Overall, seven different datasets were built based on the following criteria:
300 1. Ligand collection (D_B): This dataset contains 4548 ligands with experimental $pK_i>5$ or
301 $pIC_{50}>5$ values (i.e., <10 μM) towards both CB_1R and CB_2R. This dataset consists of all
302 selective and non-selective ligands.
303 2. Random collection (D_R): 10,000 randomly selected active compounds (excluding CB_1R
304 and CB_2R ligands).
305 3. GPCR collection (D_{GPCR}): 10,000 randomly selected G protein-coupled receptor (GPCR)
306 ligands with qualifying activity data (excluding CB_1R and CB_2R ligands).
307 4. Non-selective ligand collection (D_{MT}): 1251 non-selective ligands with experimental
308 $pK_i>5$ or $pIC_{50}>5$ values (i.e., <10 μM) towards both CB_1R and CB_2R. but with a
309 difference in potency of less than 100-fold (i.e. $\Delta p_{bind} \leq 2$).[48–50]
310 5. Selective ligand collection (D_{ST}): 3297 selective ligands with bioactivity $pK_i>5$ or
311 $pIC_{50}>5$ values (i.e., <10 μM) for both CB_1R or CB_2R, but with a difference higher than
312 100 fold (i.e. $\Delta p_{bind} >2$).
313 6. CB_1R collection (D_{CB1}): 1477 specific ligands with $pK_i>5$ or $pIC_{50}>5$ values (i.e., <10
314 μM) for CB_1R only.
315 7. CB_2R collection (D_{CB2}): 1820 specific ligands with $pK_i>5$ or $pIC_{50}>5$ values (i.e., <10
316 μM) for CB_2R only.
317 **External set compounds**

For challenging the generalization strength of the models, 1860 CB_1R and CB_2R ligands with qualifying activity values of $pEC_{50}>5$ (i.e., <10 μM) collected including 342 CB_1R and 1518 CB_2R ligands not used for model derivation.

AS analysis

From all compound datasets, AS with single or multiple substitution sites were systematically identified using the Compound-Core Relationship (CCR) method.[51] Analogs from the same series often share biological activity.[52] Nonetheless, small structural modifications might dramatically affect activity (leading to activity cliffs or inactive compounds). In our analysis, AS were used to preliminary explore structure-activity relationships (SARs) and help explain CB_1R and CB_2R ligand selectivity. Therefore, overlapping and non-overlapping core structures of AS were isolated from the sets of selective (i.e, D_{CB1}, D_{CB2}) and non-selective (i.e., D_{MT}) ligands.

Molecular representation

The recently developed Core-Substituent Fingerprint (CSFP) was employed as a molecular representation. In brief, the goal of the CSFP design was to create an easily interpretable structural fingerprint (FP) composed of molecular fragments representing as many compounds as possible with the least amount of structural information *per* molecule. CSFP comprises a total of 1000 bits with balanced composition of rings and substituents (500 instances each). Each structural fragment was assigned to a single bit position. For further details, the interested reader is referred to the original work. [21]

Machine learning models

Classification models were built by using the RF algorithm, with a default number of trees (n_estimator) equal to 400, [53] implemented using the scikit-learn python package. [54] RF was chosen as an established and robust ML approach for its transparency in parameter tuning and the ability to handle high-dimensional data. [53] The training sets represented a random sample of 70% of the compounds. Model performance was assessed using the remaining 30% of the compounds not encountered during training. Prediction results were averaged over 10 independent trials based on

alternative performance measures including accuracy, precision, recall, F1-score (F1) and Matthews correlation coefficient (MCC) [55], as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1-Score} = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (4)$$

$$\text{MCC} = \frac{(\text{TP x TN} - \text{FP x FN})}{\sqrt{((TP + FP)(TP + FN)(TN + FP)(TN + FN))}} \quad (5)$$

where TP, TN, FP, and FN stand for true positives, true negatives, false positives, and false negatives, respectively.

Model architecture

CIRCE employs a multi-layer ML classifier constituted by four binary RF models, each of which addresses a different prediction task. As shown in Figure 9, each layer is made by an independent binary RF classifier, which is nested into a sequential workflow.

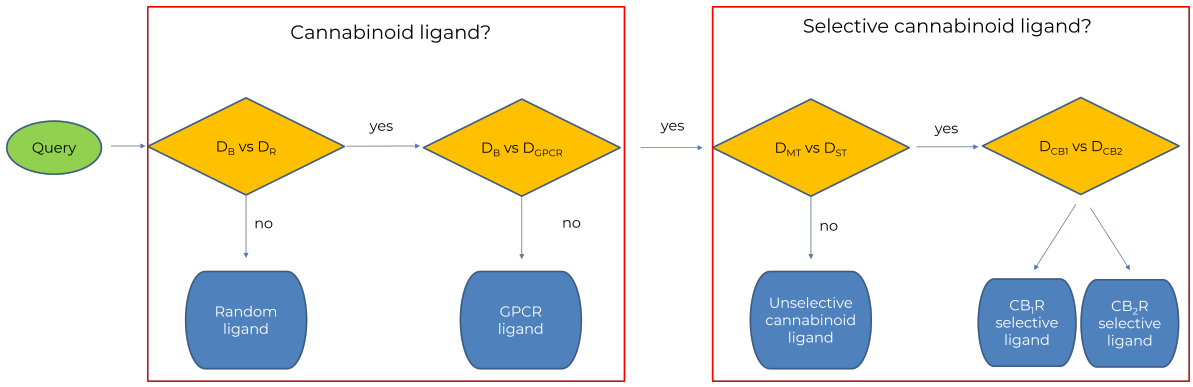


Figure 9. Sequential workflow of the multi-layer machine learning classifier.

1. D_B *vs* D_R: in the first layer, the classifier predicts the potential of a query compound to be cannabinoid receptor ligand. If the prediction is positive, the query enters the second layer (if not, not further predictions are carried out).
2. D_B *vs* D_{GPCRs}: The classifier predicts the potential of the query to preferentially bind to cannabinoid receptors compared to other GPCRs. If the prediction is positive, the query enters the third layer.
3. D_{MT} *vs* D_{ST}: The third classification distinguishes between non-selective and selective CB_1R or CB_2R ligands. If the query is predicted to be selective, it is forwarded to the fourth layer.
4. D_{CB1} *vs* D_{CB2}: The last classifier predicts a query to be a selective CB_1R or, alternatively, CB_2R ligand.

Overall, the first two layers assess the likelihood of a query compound to be considered as cannabinoid ligand. On the other hand, layers three and four predict the potential selectivity against the two Cannabinoid receptors subtypes.

Feature importance analysis

The SHAP analysis concept originated from cooperative game theory game theory.[15] In our XML analysis, fingerprint features corresponded to players engaging in the game of predicting an individual test compound. The sum of all feature importance values gives the probability of a prediction.[16,17] For RF, SHAP values were computed using the TreeExplainer algorithm.[16] To increase the transparency and the reliability of this evaluation, SHAP values were computer for each individual prediction trial and then averaged over all 10 trials.

CIRCE web platform

Both multi-layer ML classifier and the associated SHAP analysis have been implemented in a user-friendly free web platform. The web frontend of CIRCE was conceived to allow both human operation and data retrieval through POST requests. The currently available output format is an HTML page. Python Flask web framework and Jinja2 templating libraries[56] have been used for

391 building the web frontend. A graphical widget (made available by the JSME open-source project)[57]
392 enabling users to draw molecules or enter them in various input formats (SDF, MOL and InChI
393 key) is featured on the prediction interface.
394

ASSOCIATED CONTENT

Supporting Information

Figure S1: MCC values within probability range from 0.1 to 0.9 of D_B *vs* D_R, D_B *vs* D_{GPCRs}, D_{MT} *vs* D_{ST}, D_{CB1} *vs* D_{CB2} models are shown in panels (a), (b), (c) and (d), respectively. Figure S2: Tanimoto similarity computed between D_B dataset and the external test set (1860 compounds) is represented as histograms and colored in blue. Table S1: Validation studies performed by changing the threshold to 0.2 and 0.3 for the first two models. File_S1.csv, File_S2.csv, File_S3.csv, File_S4.csv, File_S5.csv, File_S6.csv and File_S7.csv report D_B, D_R, D_{GPCR}, D_{MT}, D_{ST}, D_{CB1} and D_{CB2} collections, respectively.

AUTHOR INFORMATION

Corresponding Authors

* Jürgen Bajorath − Department of Life Science Informatics and Data Science, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Friedrich-Hirzebruch-Allee 5/6, D-53115 Bonn, Germany; https://orcid.org/0000-0002-0557-5714; Email: bajorath@bit.uni-bonn.de

* Orazio Nicolotti − Department of Pharmacy-Pharmaceutical Sciences, University of the Studies of Bari “Aldo Moro”, Via E. Orabona 4, 70125, Bari, Italy; https://orcid.org/0000-0001-6533-5539; Phone: 0039 0805442551; Email: orazio.nicolotti@uniba.it.

Authors

Nicola Gambacorta − Department of Pharmacy-Pharmaceutical Sciences, University of the Studies of Bari "Aldo Moro", Via E.Orabona 4, 70125, Bari, Italy; https://orcid.org/0000-0003-1965-1519.

Fulvio Ciriaco − Department of Chemistry, University of the Studies of Bari "Aldo Moro", via E. Orabona 4, 70125, Bari, Italy; https://orcid.org/0000-0002-0695-6607.

Nicola Amoroso − Department of Pharmacy-Pharmaceutical Sciences, University of the Studies of Bari "Aldo Moro", Via E.Orabona 4, 70125, Bari, Italy; https://orcid.org/0000-0003-0211-0783.

Cosimo Damiano Altomare − Department of Pharmacy-Pharmaceutical Sciences, University of the Studies of Bari "Aldo Moro", Via E.Orabona 4, 70125, Bari, Italy; https://orcid.org/0000-0001-5016-5805.

Funding Sources

NG, FC, NA, CDA and ON gratefully thank Horizon Europe Seeds “L’intelligenza artificiale a tutela della salute in età pediatrica. Implementazione di una piattaforma digitale per il design di farmaci pediatrici sicuri”, Università degli Studi di Bari (Bari, Italy) (CUP: H99J21017390006).

Notes

The authors declare no competing financial interest.

Abbreviations Used: CB_1R, Cannabinoid Receptor type 1; CB_2R; Cannabinoid Receptor type 2; AS, analog series; CCR, compound-core relationship; CSFP, Core-Substituent Fingerprint; RF, Random Forest; ML, Machine Learning; XML, Explainable Machine Learning; SARs, structure-activity relationships; D_B, Ligands Collection; D_R, Random Collection; D_{GPCR}, GPCR Collection; D_{MT}, Non-selective Ligands; D_{ST}, Selective Ligands Collection; D_{CB1}, CB_1R collection; D_{CB2}, CB_2R collection.

References

(1) Gasperi, V.; Guzzo, T.; Topai, A.; Gambacorta, N.; Ciriaco, F.; Nicolotti, O.; Maccarrone, M. Recent Advances on Type-2 Cannabinoid (CB2) Receptor Agonists and Their Therapeutic Potential. *Curr Med Chem* **2022**. https://doi.org/10.2174/0929867329666220825161603.

(2) Catani, V. M.; Gasperi, V. Assay of CB1 Receptor Binding. In *Endocannabinoid Signaling: Methods and Protocols*; Maccarrone, M., Ed.; Methods in Molecular Biology; Springer: New York, NY, 2016; pp 41–55. https://doi.org/10.1007/978-1-4939-3539-0_5.

(3) Miranda, K.; Mehrpouya-Bahrami, P.; Nagarkatti, P. S.; Nagarkatti, M. Cannabinoid Receptor 1 Blockade Attenuates Obesity and Adipose Tissue Type 1 Inflammation Through MiR-30e-5p Regulation of Delta-Like-4 in Macrophages and Consequently Downregulation of Th1 Cells. *Frontiers in Immunology* **2019**, *10*.

(4) Rossi, F.; Punzo, F.; Umano, G. R.; Argenziano, M.; Miraglia Del Giudice, E. Role of Cannabinoids in Obesity. *International Journal of Molecular Sciences* **2018**, *19* (9), 2690. https://doi.org/10.3390/ijms19092690.

(5) Knani, I.; Earley, B. J.; Udi, S.; Nemirovski, A.; Hadar, R.; Gammal, A.; Cinar, R.; Hirsch, H. J.; Pollak, Y.; Gross, I.; Eldar-Geva, T.; Reyes-Capo, D. P.; Han, J. C.; Haqq, A. M.; Gross-Tsur, V.; Wevrick, R.; Tam, J. Targeting the Endocannabinoid/CB1 Receptor System for Treating Obesity in Prader–Willi Syndrome. *Molecular Metabolism* **2016**, *5* (12), 1187–1199. https://doi.org/10.1016/j.molmet.2016.10.004.

(6) Alenabi, A.; Malekinejad, H. Cannabinoids Pharmacological Effects Are beyond the Palliative Effects: CB2 Cannabinoid Receptor Agonist Induced Cytotoxicity and Apoptosis in Human Colorectal Cancer Cells (HT-29). *Mol Cell Biochem* **2021**, *476* (9), 3285–3301. https://doi.org/10.1007/s11010-021-04158-6.

(7) Battista, N.; Di Tommaso, M.; Bari, M.; Maccarrone, M. The Endocannabinoid System: An Overview. *Frontiers in Behavioral Neuroscience* **2012**, *6*, 9. https://doi.org/10.3389/fnbeh.2012.00009.

(8) Cassano, T.; Calcagnini, S.; Pace, L.; De Marco, F.; Romano, A.; Gaetani, S. Cannabinoid Receptor 2 Signaling in Neurodegenerative Disorders: From Pathogenesis to a Promising Therapeutic Target. *Frontiers in Neuroscience* **2017**, *11*.

(9) Caffarel, M. M.; Andradas, C.; Mira, E.; Pérez-Gómez, E.; Cerutti, C.; Moreno-Bueno, G.; Flores, J. M.; García-Real, I.; Palacios, J.; Mañes, S.; Guzmán, M.; Sánchez, C. Cannabinoids Reduce ErbB2-Driven Breast Cancer Progression through Akt Inhibition. *Molecular Cancer* **2010**, *9* (1), 196. https://doi.org/10.1186/1476-4598-9-196.

(10) Brennecke, B.; Gazzi, T.; Atz, K.; Fingerle, J.; Kuner, P.; Schindler, T.; Weck, G. de; Nazaré, M.; Grether, U. Cannabinoid Receptor Type 2 Ligands: An Analysis of Granted Patents since 2010. *Pharmaceutical Patent Analyst* **2021**, *10* (3), 111–163. https://doi.org/10.4155/ppa-2021-0002.

(11) Busquets Garcia, A.; Soria-Gomez, E.; Bellocchio, L.; Marsicano, G. Cannabinoid Receptor Type-1: Breaking the Dogmas. *F1000Res* **2016**, *5*, 990. https://doi.org/10.12688/f1000research.8245.1.

(12) Badal, S.; Smith, K. N.; Rajnarayanan, R. Analysis of Natural Product Regulation of Cannabinoid Receptors in the Treatment of Human Disease. *Pharmacology & Therapeutics* **2017**, *180*, 24–48. https://doi.org/10.1016/j.pharmthera.2017.06.003.

(13) Fattore, L.; Fratta, W. Beyond THC: The New Generation of Cannabinoid Designer Drugs. *Frontiers in Behavioral Neuroscience* **2011**, *5*.

(14) Palmer, S. L.; Thakur, G. A.; Makriyannis, A. Cannabinergic Ligands. *Chemistry and Physics of Lipids* **2002**, *121* (1), 3–19. https://doi.org/10.1016/S0009-3084(02)00143-3.

(15) Hourani, W.; Alexander, S. P. H. Cannabinoid Ligands, Receptors and Enzymes: Pharmacological Tools and Therapeutic Potential. *Brain and Neuroscience Advances* **2018**, *2*, 239821281878390. https://doi.org/10.1177/2398212818783908.

(16) Dara, S.; Dhamercherla, S.; Jadav, S. S.; Babu, C. M.; Ahsan, M. J. Machine Learning in Drug Discovery: A Review. *Artif Intell Rev* **2022**, *55* (3), 1947–1999. https://doi.org/10.1007/s10462-021-10058-4.
(17) Lavecchia, A. Machine-Learning Approaches in Drug Discovery: Methods and Applications. *Drug Discovery Today* **2015**, *20* (3), 318–331. https://doi.org/10.1016/j.drudis.2014.10.012.
(18) Lamens, A.; Bajorath, J. Explaining Accurate Predictions of Multitarget Compounds with Machine Learning Models Derived for Individual Targets. *Molecules* **2023**, *28* (2), 825. https://doi.org/10.3390/molecules28020825.
(19) Feldmann, C.; Bajorath, J. Calculation of Exact Shapley Values for Support Vector Machines with Tanimoto Kernel Enables Model Interpretation. *iScience* **2022**, *25* (9), 105023. https://doi.org/10.1016/j.isci.2022.105023.
(20) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; Hersey, A.; Leach, A. R. ChEMBL: Towards Direct Deposition of Bioassay Data. *Nucleic Acids Res* **2019**, *47* (D1), D930–D940. https://doi.org/10.1093/nar/gky1075.
(21) Janela, T.; Takeuchi, K.; Bajorath, J. Introducing a Chemically Intuitive Core-Substituent Fingerprint Designed to Explore Structural Requirements for Effective Similarity Searching and Machine Learning. *Molecules* **2022**, *27* (7), 2331. https://doi.org/10.3390/molecules27072331.
(22) Merrick, L.; Taly, A. The Explanation Game: Explaining Machine Learning Models Using Shapley Values. In *Machine Learning and Knowledge Extraction*; Holzinger, A., Kieseberg, P., Tjoa, A. M., Weippl, E., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, 2020; pp 17–38. https://doi.org/10.1007/978-3-030-57321-8_2.
(23) Rodríguez-Pérez, R.; Bajorath, J. Interpretation of Machine Learning Models Using Shapley Values: Application to Compound Potency and Multi-Target Activity Predictions. *J Comput Aided Mol Des* **2020**, *34* (10), 1013–1026. https://doi.org/10.1007/s10822-020-00314-0.
(24) Togo, M. V.; Mastrolorito, F.; Ciriaco, F.; Trisciuzzi, D.; Tondo, A. R.; Gambacorta, N.; Bellantuono, L.; Monaco, A.; Leonetti, F.; Bellotti, R.; Altomare, C. D.; Amoroso, N.; Nicolotti, O. TIRESIA: An EXplainable Artificial Intelligence Platform for Predicting Developmental Toxicity. *J. Chem. Inf. Model.* **2023**, *63* (1), 56–66. https://doi.org/10.1021/acs.jcim.2c01126.
(25) Jeong, M.; Nam, J.; Ko, B. C. Lightweight Multilayer Random Forests for Monitoring Driver Emotional Status. *IEEE Access* **2020**, *8*, 60344–60354. https://doi.org/10.1109/ACCESS.2020.2983202.
(26) Ciriaco, F.; Gambacorta, N.; Trisciuzzi, D.; Nicolotti, O. PLATO: A Predictive Drug Discovery Web Platform for Efficient Target Fishing and Bioactivity Profiling of Small Molecules. *International Journal of Molecular Sciences* **2022**, *23* (9), 5245. https://doi.org/10.3390/ijms23095245.
(27) Ciriaco, F.; Gambacorta, N.; Alberga, D.; Nicolotti, O. Quantitative Polypharmacology Profiling Based on a Multifingerprint Similarity Predictive Approach. *J. Chem. Inf. Model.* **2021**, *61* (10), 4868–4876. https://doi.org/10.1021/acs.jcim.1c00498.
(28) Ciriaco, F.; Gambacorta, N.; Leonetti, F.; Altomare, C. D.; Nicolotti, O. Virtual Reverse Screening Approach to Target Type 2 Cannabinoid Receptor. *Methods Mol Biol* **2023**, *2576*, 495–504. https://doi.org/10.1007/978-1-0716-2728-0_40.
(29) Aviz-Amador, A.; Contreras-Puentes, N.; Mercado-Camargo, J. Virtual Screening Using Docking and Molecular Dynamics of Cannabinoid Analogs against CB1 and CB2 Receptors. *Computational Biology and Chemistry* **2021**, *95*, 107590. https://doi.org/10.1016/j.compbiolchem.2021.107590.

(30) Mizera, M.; Latek, D.; Cielecka-Piontek, J. Virtual Screening of C. Sativa Constituents for the Identification of Selective Ligands for Cannabinoid Receptor 2. *International Journal of Molecular Sciences* **2020**, *21* (15), 5308. https://doi.org/10.3390/ijms21155308.
(31) Uba, A. I.; Aluwala, H.; Liu, H.; Wu, C. Elucidation of Partial Activation of Cannabinoid Receptor Type 2 and Identification of Potential Partial Agonists: Molecular Dynamics Simulation and Structure-Based Virtual Screening. *Computational Biology and Chemistry* **2022**, *99*, 107723. https://doi.org/10.1016/j.compbiolchem.2022.107723.
(32) Hanachi, R.; Said, R. B.; Allal, H.; Rahali, S.; Alkhalifah, M. A. M.; Alresheedi, F.; Tangour, B.; Hochlaf, M. Structural, QSAR, Machine Learning and Molecular Docking Studies of 5-Thiophen-2-Yl Pyrazole Derivatives as Potent and Selective Cannabinoid-1 Receptor Antagonists. *New J. Chem.* **2021**, *45* (38), 17796–17807. https://doi.org/10.1039/D1NJ02261J.
(33) Burlacu, C. M.; Praisler, M.; Burlacu, A. C. Computerized Detection of JWH Synthetic Cannabinoids Class Membership Based on Machine Learning Algorithms and Molecular Descriptors. In *2022 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*; 2022; pp 1–5. https://doi.org/10.1109/AQTR55203.2022.9801971.
(34) Zhou, H.; Shan, M.; Qin, L.-P.; Cheng, G. Reliable Prediction of Cannabinoid Receptor 2 Ligand by Machine Learning Based on Combined Fingerprints. *Computers in Biology and Medicine* **2023**, *152*, 106379. https://doi.org/10.1016/j.compbiomed.2022.106379.
(35) Atz, K.; Guba, W.; Grether, U.; Schneider, G. Machine Learning and Computational Chemistry for the Endocannabinoid System. In *Endocannabinoid Signaling: Methods and Protocols*; Maccarrone, M., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2023; pp 477–493. https://doi.org/10.1007/978-1-0716-2728-0_39.
(36) El-Atawneh, S.; Hirsch, S.; Hadar, R.; Tam, J.; Goldblum, A. Prediction and Experimental Confirmation of Novel Peripheral Cannabinoid-1 Receptor Antagonists. *J. Chem. Inf. Model.* **2019**, *59* (9), 3996–4006. https://doi.org/10.1021/acs.jcim.9b00577.
(37) Rodrigues, T.; Bernardes, G. J. L. Machine Learning for Target Discovery in Drug Development. *Current Opinion in Chemical Biology* **2020**, *56*, 16–22. https://doi.org/10.1016/j.cbpa.2019.10.003.
(38) Gambacorta, N.; Gasperi, V.; Guzzo, T.; Di Leva, F. S.; Ciriaco, F.; Sánchez, C.; Tullio, V.; Rozzi, D.; Marinelli, L.; Topai, A.; Nicolotti, O.; Maccarrone, M. Exploring the 1,3-Benzoxazine Chemotype for Cannabinoid Receptor 2 as a Promising Anti-Cancer Therapeutic. *European Journal of Medicinal Chemistry* **2023**, *259*, 115647. https://doi.org/10.1016/j.ejmech.2023.115647.
(39) Jia, A.; Xu, L.; Wang, Y. Venn Diagrams in Bioinformatics. *Briefings in Bioinformatics* **2021**, *22* (5), bbab108. https://doi.org/10.1093/bib/bbab108.
(40) Leite, C. E.; Mocelin, C. A.; Petersen, G. O.; Leal, M. B.; Thiesen, F. V. Rimonabant: An Antagonist Drug of the Endocannabinoid System for the Treatment of Obesity. *Pharmacological Reports* **2009**, *61* (2), 217–224. https://doi.org/10.1016/S1734-1140(09)70025-8.
(41) Pi-Sunyer, F. X.; Aronne, L. J.; Heshmati, H. M.; Devin, J.; Rosenstock, J.; RIO-North America Study Group, for the. Effect of Rimonabant, a Cannabinoid-1 Receptor Blocker, on Weight and Cardiometabolic Risk Factors in Overweight or Obese PatientsRIO-North America: A Randomized Controlled Trial. *JAMA* **2006**, *295* (7), 761–775. https://doi.org/10.1001/jama.295.7.761.
(42) Haspula, D.; Clark, M. A. Cannabinoid Receptors: An Update on Cell Signaling, Pathophysiological Roles and Therapeutic Opportunities in Neurological, Cardiovascular, and Inflammatory Diseases. *International Journal of Molecular Sciences* **2020**, *21* (20), 7693. https://doi.org/10.3390/ijms21207693.
(43) Li, X.; Hua, T.; Vemuri, K.; Ho, J.-H.; Wu, Y.; Wu, L.; Popov, P.; Benchama, O.; Zvonok, N.; Locke, K.; Qu, L.; Han, G. W.; Iyer, M. R.; Cinar, R.; Coffey, N. J.; Wang, J.; Wu, M.; Katritch, V.; Zhao, S.; Kunos, G.; Bohn, L. M.; Makriyannis, A.; Stevens, R. C.; Liu, Z.-J.

Crystal Structure of the Human Cannabinoid Receptor CB2. *Cell* **2019**, *176* (3), 459-467.e13. https://doi.org/10.1016/j.cell.2018.12.011.
(44) Fulo, H. F.; Shoeib, A.; Cabanlong, C. V.; Williams, A. H.; Zhan, C.-G.; Prather, P. L.; Dudley, G. B. Synthesis, Molecular Pharmacology, and Structure–Activity Relationships of 3-(Indanoyl)Indoles as Selective Cannabinoid Type 2 Receptor Antagonists. *J. Med. Chem.* **2021**, *64* (9), 6381–6396. https://doi.org/10.1021/acs.jmedchem.1c00442.
(45) Iyer, M. R.; Cinar, R.; Wood, C. M.; Zawatsky, C. N.; Coffey, N. J.; Kim, K. A.; Liu, Z.; Katz, A.; Abdalla, J.; Hassan, S. A.; Lee, Y.-S. Synthesis, Biological Evaluation, and Molecular Modeling Studies of 3,4-Diarylpyrazoline Series of Compounds as Potent, Nonbrain Penetrant Antagonists of Cannabinoid-1 (CB_1 R) Receptor with Reduced Lipophilicity. *J. Med. Chem.* **2022**, *65* (3), 2374–2387. https://doi.org/10.1021/acs.jmedchem.1c01836.
(46) Alberga, D.; Trisciuzzi, D.; Montaruli, M.; Leonetti, F.; Mangiatordi, G. F.; Nicolotti, O. A New Approach for Drug Target and Bioactivity Prediction: The Multifingerprint Similarity Search Algorithm (MuSSeL). *J Chem Inf Model* **2019**, *59* (1), 586–596. https://doi.org/10.1021/acs.jcim.8b00698.
(47) Montaruli, M.; Alberga, D.; Ciriaco, F.; Trisciuzzi, D.; Tondo, A. R.; Mangiatordi, G. F.; Nicolotti, O. Accelerating Drug Discovery by Early Protein Drug Target Prediction Based on a Multi-Fingerprint Similarity Search †. *Molecules* **2019**, *24* (12), 2233. https://doi.org/10.3390/molecules24122233.
(48) Bajorath, J. Explainable Machine Learning for Medicinal Chemistry: Exploring Multi-Target Compounds. *Future Medicinal Chemistry* **2022**, fmc-2022-0122. https://doi.org/10.4155/fmc-2022-0122.
(49) Feldmann, C.; Yonchev, D.; Bajorath, J. Structured Data Sets of Compounds with Multi-Target and Corresponding Single-Target Activity from Biological Assays. *Future Science OA* **2021**, *7* (5), FSO685. https://doi.org/10.2144/fsoa-2020-0209.
(50) Feldmann, C.; Bajorath, J. Differentiating Inhibitors of Closely Related Protein Kinases with Single- or Multi-Target Activity via Explainable Machine Learning and Feature Analysis. *Biomolecules* **2022**, *12* (4), 557. https://doi.org/10.3390/biom12040557.
(51) Naveja, J. J.; Vogt, M.; Stumpfe, D.; Medina-Franco, J. L.; Bajorath, J. Systematic Extraction of Analogue Series from Large Compound Collections Using a New Computational Compound–Core Relationship Method. *ACS Omega* **2019**, *4* (1), 1027–1032. https://doi.org/10.1021/acsomega.8b03390.
(52) Cruz-Monteagudo, M.; Medina-Franco, J. L.; Pérez-Castillo, Y.; Nicolotti, O.; Cordeiro, M. N. D. S.; Borges, F. Activity Cliffs in Drug Discovery: Dr Jekyll or Mr Hyde? *Drug Discovery Today* **2014**, *19* (8), 1069–1080. https://doi.org/10.1016/j.drudis.2014.02.003.
(53) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1947–1958. https://doi.org/10.1021/ci034160g.
(54) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. arXiv June 5, 2018. https://doi.org/10.48550/arXiv.1201.0490.
(55) Korotcov, A.; Tkachenko, V.; Russo, D. P.; Ekins, S. Comparison of Deep Learning With Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets. *Mol. Pharmaceutics* **2017**, *14* (12), 4462–4475. https://doi.org/10.1021/acs.molpharmaceut.7b00578.
(56) Grinberg, M. (2018). Flask Web Development: Developing Web Applications with Python.
(57) Bienfait, B.; Ertl, P. JSME: A Free Molecule Editor in JavaScript. *Journal of Cheminformatics* **2013**, *5* (1), 24. https://doi.org/10.1186/1758-2946-5-24.

638

Table of Contents Graphic

CIRCE: Web-Based Platform for the Prediction of Cannabinoid Receptor Ligands Using Explainable Machine Learning

Nicola Gambacorta, Fulvio Ciriaco, Nicola Amoroso, Cosimo Damiano Altomare, Jürgen Bajorath and Orazio Nicolotti