

Optimization of metabolic stability of ligands of serotonin receptor 5-HT7 using SHAP values

Agnieszka Pocha¹, Rafał Jankowski¹, Sabina Podlewska^{2,3*}

¹Faculty of Mathematics and Computer Science, Jagiellonian University, 6 S. Łojasiewicza Street, 30-348 Kraków, Poland.

²Maj Institute of Pharmacology Polish Academy of Sciences, 12 Smętna Street, 31-343 Kraków, Poland

³ Department of Technology and Biotechnology of Drugs, Jagiellonian University, Medical College, Medyczna Street 9, 30-688 Cracow, Poland

*e-mail: smusz@if-pan.krakow.pl

Background

The process of drug design and development is very complex and involves multiple stages. After detection of compounds active towards considered target, scientists must also optimize its physicochemical and pharmacokinetic properties, so as it reaches the desirable therapeutic area, it does not undergo fast decomposition and it does not produce toxic effects. In the study, we focused on metabolic stability – very important parameter influencing the compound potential to become a future drug. After entering organism, the drug undergoes a number of processes, which might lead to modification of its structure, and therefore, evaluation of metabolic stability *in silico* is very complex and difficult task.

Aim of the study

The aim of the project is to construct a web-based service for evaluation of metabolic stability of compounds with the special focus on the assessment of the structural contributions to particular predictions. It was obtained via the application of SHapley Additive exPlanations (SHAP) values. The service is available at <https://metstab-shap.matinf.uj.edu.pl/>

Results

The workflow of the developed methodology is presented in Figure 1. It involves preparation of the dataset on the basis of information present in the ChEMBL database, representation of compounds with the use of two substructural fingerprints: MACCSFP and Klekota&Roth Fingerprint (KRFP), development of machine learning-based predictive models, and calculation of SHAP values.

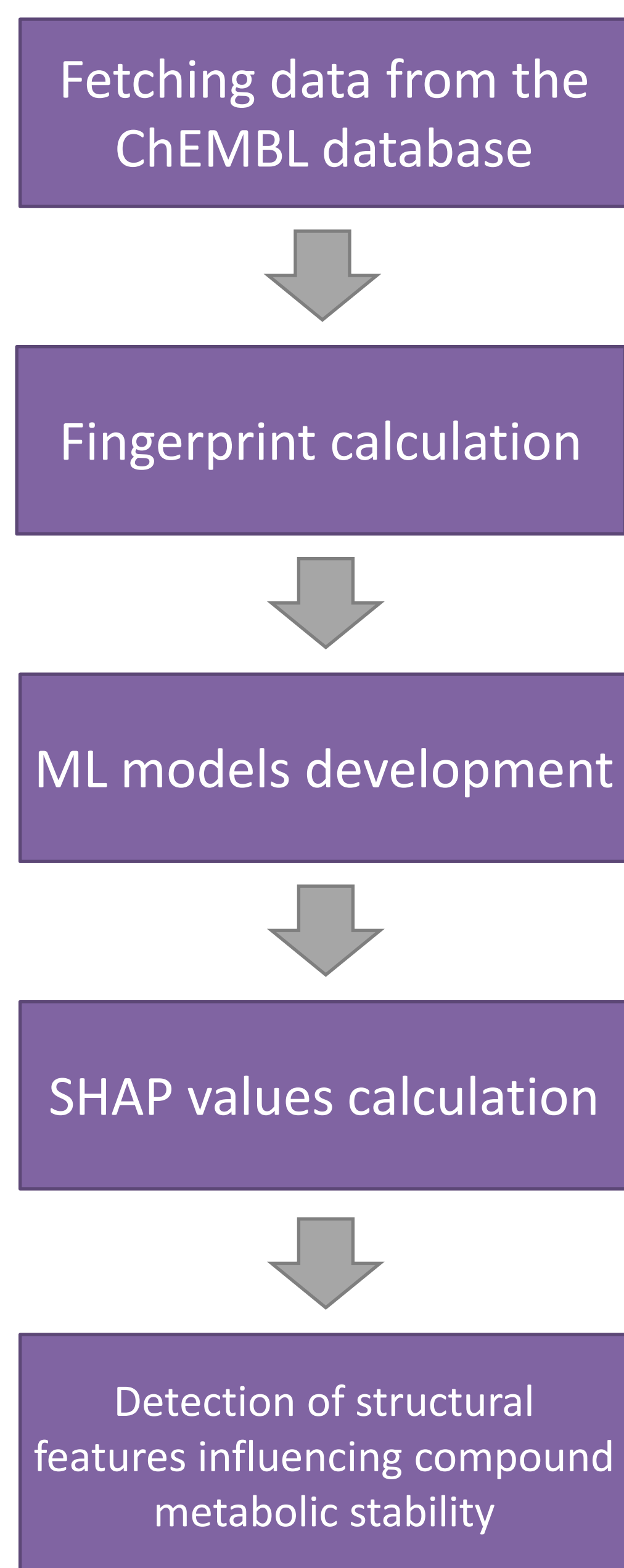


Figure 1. Scheme of the developed protocol for optimization of compound metabolic stability

The machine learning algorithms used in the study were Naive Bayes, Support Vector Machines and trees.

Analysis of the ChEMBL dataset

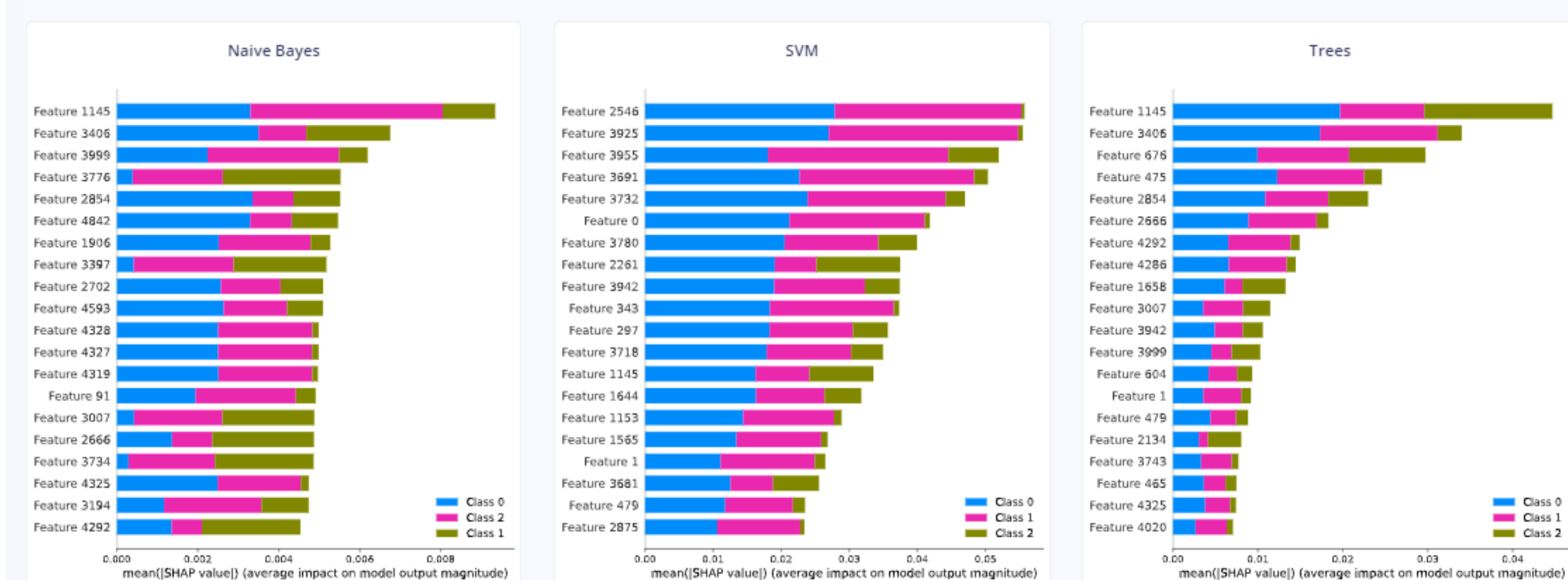
Results of analysis of the ChEMBL datasets with the use of the following approaches: Naive Bayes, SVMs and trees. The following half-lifetime cut-offs are used for classification studies:

- ≤ 0.6 – low stability (Class 0)
- $(0.6-2.32)$ – medium stability (Class 1)
- 2.32 – high stability (Class 2)

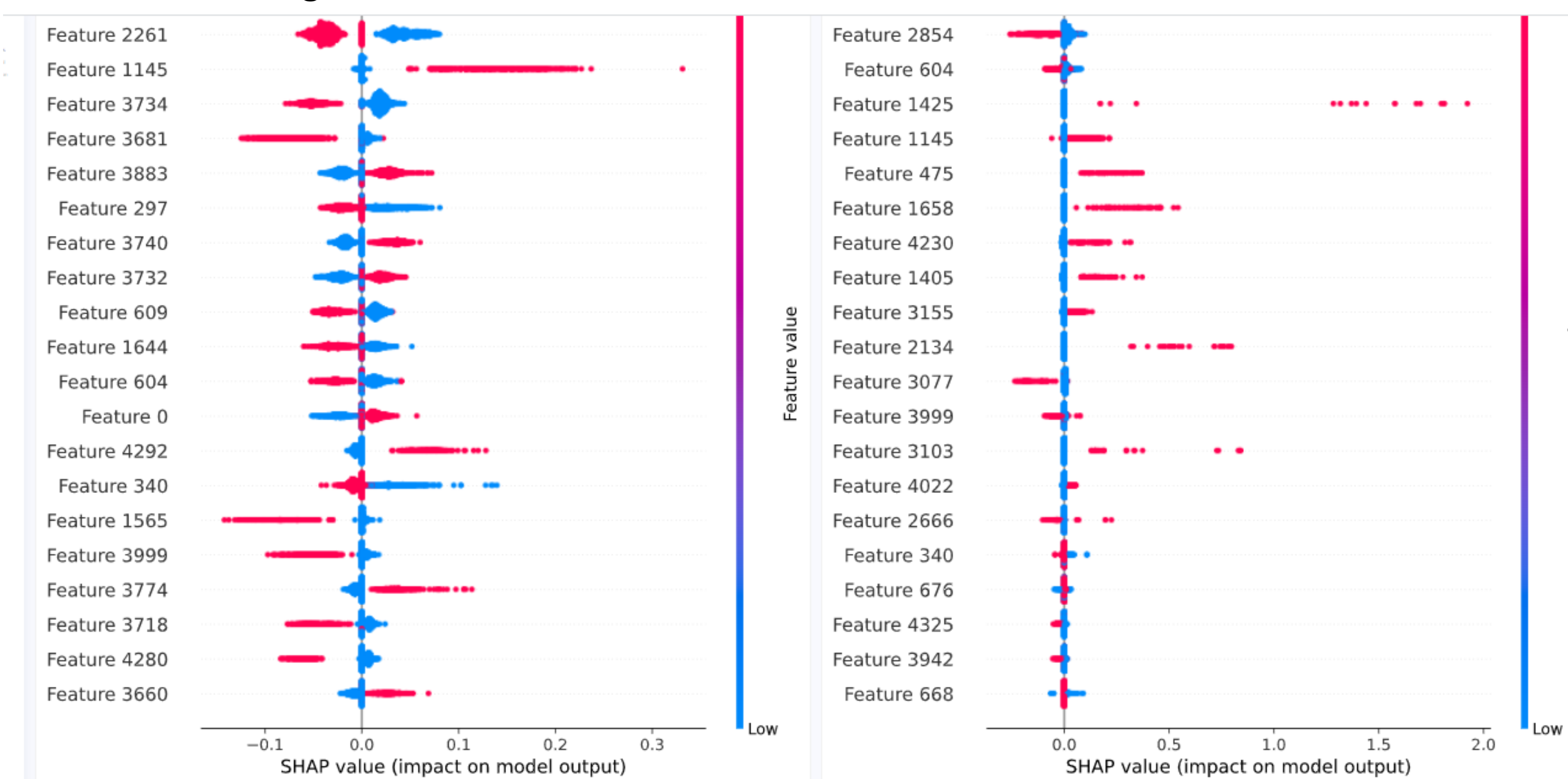
The compounds were represented with the use of two substructural fingerprints: MACCSFP and Klekota-Roth Fingerprint. SHapley Additive exPlanations (SHAP) were used to determine the influence of particular structural features on the predictions of the models. SHAP allows to attribute a single value for each feature of the input compound. This value can be interpreted as feature importance and reflects the feature's influence on the prediction.

The figures below present 20 features with the highest mean absolute SHAP value determined for each combination of dataset (human or rat), compound representation (MACCSFP or KRFP) and ML approach (Naive Bayes, SVMs, trees). For classification models, the features are plotted side by side starting from the actual prediction and the most important feature at the top. The SHAP values of the remaining features are summed and plotted collectively at the bottom of the plot and ending at the model's average prediction. In case of regression, each dot represents a single correct prediction, its color the value of the corresponding feature (blue: absence, red: presence), and the position on the x-axis is the SHAP value itself. Keys referring to particular feature numbers are explained in the "Features description" tab.

Human - KRFP - Classification



Human-KRFP-Regression



Analyze custom compound

Here, you can analyze the metabolic stability of your compound and examine the influence of particular chemical moieties. Please, submit your compound in the SMILES format (maximum number of 5 compounds can be submitted in one run) and choose a model and compound representation. In addition to analysis for the submitted structures, SHAP values for the most similar compound from the ChEMBL dataset are determined and visualized (the most similar structure is returned only when the Tanimoto coefficient calculated on Morgan fingerprints is higher than 0.7). It can help with the optimization of the metabolic stability of your compound.

Custom compound submission

SMILES

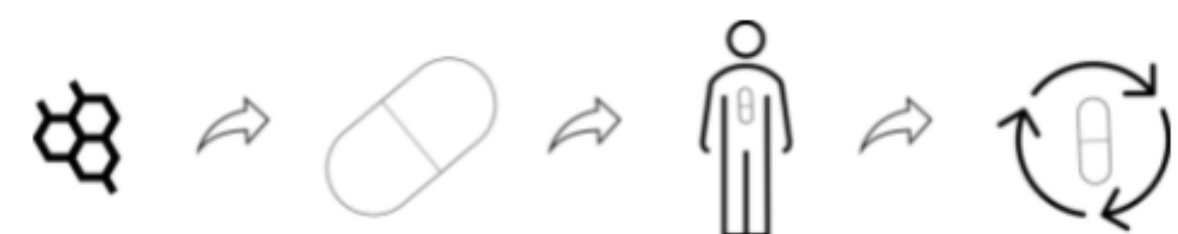
CC(=O)N1CC(N)(C)C1c1ccc(Cl)cc1Cl

Model and representation

Human - MACCS - Classification

Evaluation of structural contributions to metabolic stability predictions with the use of SHAP values

Introduction



Metabolic stability is an important parameter influencing the compound chances to be a future drug, as it determines the time that the compound can act in the organism and play its role as a drug. Due to great complexity of xenobiotic transformation pathways in living organisms, the evaluation and optimization of metabolic stability remains a big challenge.

This webservice can help in the evaluation and analysis of structural features influencing metabolic stability. It is based on the determination of SHAP values for several predictive models. The service allows a detailed analysis of SHAP values obtained for ChEMBL datasets, as well as their determination and analysis for any compound submitted by the user.

Output of the presented tool can be of great help in the design of new ligands with improved metabolic stability, helping in the detection of privileged and unfavourable chemical moieties during stability optimization.

Figure 2. Screen of the main page of the prepared web service.

Acknowledgments

The study was supported by the grant OPUS 2018/31/B/NZ2/00165 financed by the National Science Centre, Poland (www.ncn.gov.pl).