Machine-learning-based module for the design of polypharmacological drugs

Sabina Podlewska, Rafał Kurczab

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

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

The simplest way of searching for a new biologically active substance is to search for a drug that will bind with a particular biological target. However, hits found within such a procedure very often fail at the subsequent stages of the drug design pipeline, due to the adverse effects caused by them, which is most often related to the interaction with the so-called off-targets. Therefore, nowadays, the most advanced procedures consider multiple targets at the same time, following the concept of polypharmacology.

In the study, we developed a polypharmacological machine-learning-based module for virtual screening. It is part of a multi-tool platform, enabling comprehensive approach to solve polypharmacological tasks. The module allows evaluation of compound libraries towards almost 200 protein targets, out of which several (selected by the user) can be consider simultaneously. The ML algorithms were trained and optimized for each target separately on the data present in the ChEMBL database. For compound representation, three fingerprints were used: one of a hashed typeand two substructural ones.

The module is part of the SilicoPharm polypharmacological platform (Figure 1)

Figure 1. Main page of the SilicoPharm polypharmacological platform.

CASE STUDY - OPIOID RECEPTORS

The global efficiency of the prediction power of the developed models was evaluated via the calculation of the accuracy parameter. However, we were interested not only in the assessment of the global efficiency, but also in analyzing particular case studies to indicate the possible sources of prediction errors in order to eliminate as many of tchem as possible or recommend particular compound sets to be evaluated by other methods, such as docking.

As a case study, we selected opioid receptors. All affinity values (expressed via K_i) referring to mu, kappa, and delta-opioid re-ceptors were collected. The compound structures were transformed to the bit-string representation using PaDEL descriptor software (the following fingerprints were used: Extended Fingerprint (ExtFP), Klekota&Roth Fingerprint (KlekFP) and MACCS Fingerprint (MACCSFP). K_i values were predicted using the k-nearest neighbor algorithm (IBk) and Random Forest (RF). K_i values were predicted in regression experiments. Predictions were carried out in the 10-fold cross-validation mode [1].

RESULTS – DISTRIBUTION OF PREDICTION ERROR

Analysis of the global prediction efficiency involved examination of the relative absolute error values (Figure 2).

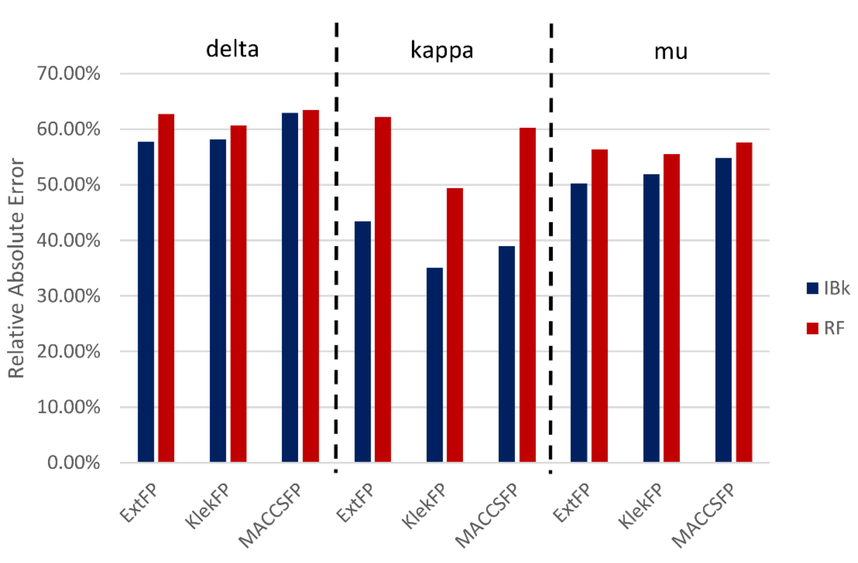

Figure 2. Relative absolute errors were obtained in the predictions of compounds Ki values for different ML algorithms and compound representations..

RESULTS

In general, RF algorithm appeared to be a little bit less accurate than predictions provided by IBk (in-dicated by higher values of Relative Absolute Error). Moreover, for the kappa-opioid receptor, the differences between results obtained by RF and IBk are the highest and equal to about 20%. Although on average the best predictions were obtained for KlekFP, only for the kappa-opioid receptor, the difference between various compound representations is not strongly indicated. When it comes to the analysis of a particular receptor subtype, in general, the most accurate results were obtained for the kappa-opioid receptor, with values of Relative Absolute Error not exceeding 50% for all compounds representations for IBk, and between 50% to 60% for RF. On the other hand, the lowest prediction power was observed for the delta-opioid receptor, where Relative Absolute Error values were around 60% for all ML methods and compound representations used, the highest for MACCSFP, with an only slight difference between IBk and RF.

ANALYSIS OF COMPOUNDS WITH THE HIGHEST PREDICTION ERROR

Venn diagrams presenting the number of overlapping compounds for each experimental settings were prepared (200 top compounds were considered in each case, Figure 3) to analyze whether the compounds for which the ML methods are unable to produce correct predictions are the same for different methods/compound representations. The highest number of compounds consistently incorrectly predicted for both representations and ML methods considered occurred for the delta-opioid receptor (82 compounds). It was similar to the number of wrongly predicted ligands from the set of mu-opioid receptor (71), whereas for the kappa-opioid receptor, the number of ligands, which were incorrectly predicted in all experimental conditions was much lower and it was equal to 35.

The total numer of overlapping compounds is rather high (25-60%); however, for compounds, which were not wrongly predicted by all representations/methods; the correct prediction can be made by shifting to other predictive approaches. Therefore, it is important during application of ML algorithms, that various combinations of methods and compound representations are used to maximize the reliability of the obtained evaluations.

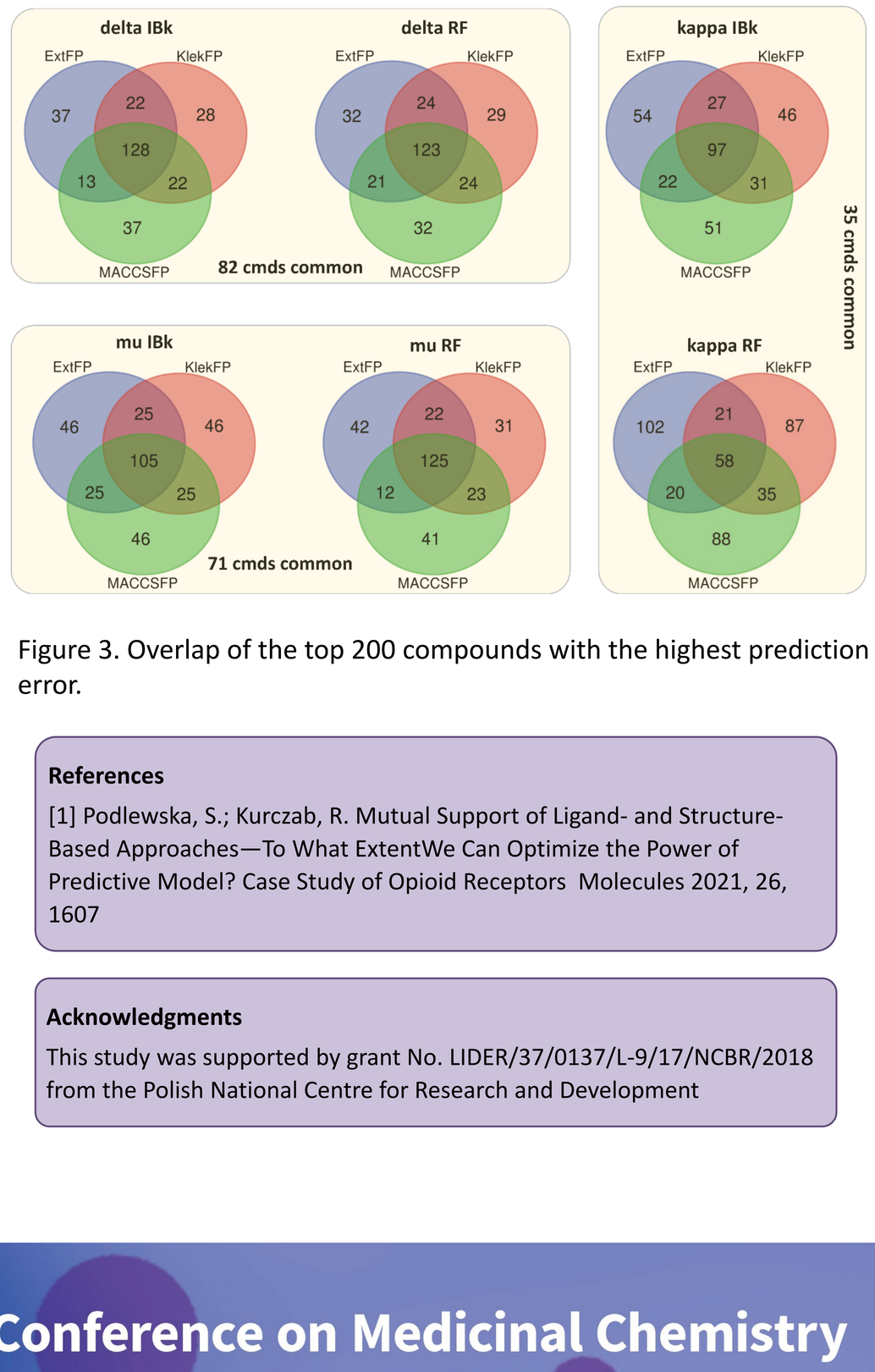

Figure 3. Overlap of the top 200 compounds with the highest prediction error.

References

[1] Podlewska, S.; Kurczab, R. Mutual Support of Ligand- and Structure-Based Approaches—To What ExtentWe Can Optimize the Power of Predictive Model? Case Study of Opioid Receptors Molecules 2021, 26, 1607

Acknowledgments

This study was supported by grant No. LIDER/37/0137/L-9/17/NCBR/2018 from the Polish National Centre for Research and Development