

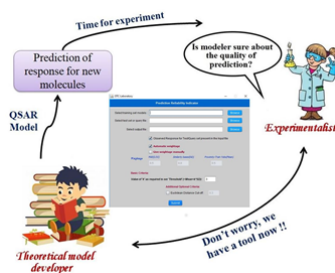
## “Prediction Reliability Indicator” : A new tool to judge the quality of predictions from QSAR models for new query compounds

Kunal Roy ([kunal.roy@jadavpuruniversity.in](mailto:kunal.roy@jadavpuruniversity.in))<sup>a</sup>, Pravin Ambure ([ambure.pharmait@gmail.com](mailto:ambure.pharmait@gmail.com))<sup>a</sup> and Supratik Kar ([karsupratik86@gmail.com](mailto:karsupratik86@gmail.com))<sup>b</sup>

<sup>a</sup>Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India

<sup>b</sup>Interdisciplinary Center for Nanotoxicity, Department of Chemistry, Physics and Atmospheric Sciences, Jackson State University, Jackson, MS-39217, USA.

### Graphical Abstract



### Abstract

Prediction of an endpoint for new query chemical without having any experimental response data is one of the important applications of Quantitative structure-activity relationship (QSAR) models. Usually a QSAR model is developed based on chemical information of a properly designed training set and corresponding experimental response data while the model is validated using one or more test set(s) for which the experimental response data are available. However, it is interesting to estimate the reliability of predictions when the model is applied to a completely new data set (true external set) even when the new data points are within applicability domain (AD) of the developed model. In the present study, we have developed a tool “Prediction Reliability Indicator” to indicate or categorize the quality of predictions for the test set or true external set into three groups: good (with composite score 3), moderate (with composite score 2) and bad (with composite score 1). Here, we have used three criteria [1) Mean absolute error of leave-one-out predictions for 10 most close training compounds for each query molecule (J Chemom 2018, <http://dx.doi.org/10.1002/cem.2992> ); 2) Applicability domain in terms of similarity based on the standardization approach (Chemom Intell Lab Sys, 145, 2015, 22-29, <http://dx.doi.org/10.1016/j.chemolab.2015.04.013>); 3) Proximity of the predicted value of the query compound to the experimental mean training response (Chemom Intell Lab Sys, 162, 2017, 44-54, <https://authors.elsevier.com/a/1UOpFcc6LvBdv> )] in different weightage schemes for making a composite

score of predictions. The tool can automatically find the optimum weightage based on % correct prediction score computed using a test set with known observed response and thus known quality of predictions. However, the user also has an option to select the weightage manually. It was found that using the most frequently appearing weightage scheme 0.5:0:0.5, the composite score based categorization showed concordance with absolute prediction error based categorization for more than 80% test data points while working with 5 different data sets with 15 models for each set derived in three different splitting techniques. These observations were also confirmed with two external sets suggesting applicability of the scheme to judge the reliability of predictions for new data sets. The tool is available free of charge at <http://dtclab.webs.com/software-tools>.

## References

1. Roy K, Ambure P, Kar S, Ojha PK, Is it possible to improve the quality of predictions from an “intelligent” use of multiple QSAR/QSPR/QSTR models? J Chemom 2018, <http://dx.doi.org/10.1002/cem.2992>
2. Roy K, Kar S, Ambure P, On a simple approach for determining applicability domain of QSAR models Chemom Intell Lab Sys, 145, 2015, 22-29, <http://dx.doi.org/10.1016/j.chemolab.2015.04.013>
3. Roy K, Ambure P, Aher R, How important is to detect systematic error in predictions and understand statistical applicability domain of QSAR models? Chemom Intell Lab Sys, 162, 2017, 44-54, <https://authors.elsevier.com/a/1UOpFcc6LvBdv>
4. <http://dtclab.webs.com/software-tools>