**Beware of external validation! – A Comparative Study of Several Validation Techniques used in QSAR Modelling**

Subhabrata Majumdar 1, Subhash C. Basak 2,\*

1 University of Florida Informatics Institute, 432 Newell Drive, CISE Bldg E251, Gainesville, FL 32611, USA; smajumdar@ufl.edu

2 Natural Resources Research Institute, University of Minnesota Duluth, and Departement of Chemistry and Biochemistry, University of Minnesota Duluth, 5013 Miller Trunk Highway,

Duluth, MN 55811, USA; sbasak@nrri.umn.edu

**\*** Correspondence: sbasak@nrri.umn.edu; Tel.: +1-218-727-1335; Fax: +1-218-720-4328

**Abstract:**

***Background***:Proper validation is an important aspect of QSAR modelling. External validation is one of the widely used validation methods in QSAR where the model is built on a subset of the data and validated on the rest of the samples. However, its effectiveness for datasets with a small number of samples but large number of predictors remains suspect.

***Objective***: Calculating hundreds or thousands of molecular descriptors using currently available software has become the norm in QSAR research, owing to computational advances in the past few decades. Thus, for *n* chemical compounds and *p* descriptors calculated for each molecule, the typical chemometric dataset today has high value of *p* but small *n* (i.e. *n << p*). Motivated by the evidence of inadequacies of external validation in estimating the true predictive capability of a statistical model in recent literature, this paper performs a comparative study of this method with several other validation techniques.

***Methodology***: We compared four validation methods: leave-one-out, *K*-fold, external and multi-split validation, using statistical models built using the LASSO regression, which simultaneously performs variable selection and modelling. We used three simulated datasets and one real dataset of 95 congeneric amine mutagens for this evaluation.

***Results***: For all the datasets, LOO has the overall best performance among all validation methods applied. External validation metrics have high variation among different random splits of the data.

***Conclusion***: Results from external validation are too unstable for the datasets we analyzed. Based on our findings, we recommend using the LOO procedure for validating QSAR models built on high-dimensional small-sample data.

Keywords: Cross validation; leave one out (LOO) cross validation; *K*-fold cross validation; external validation; LASSO; chemical mutagens; aromatic and heteroaromatic amines