Int. J. Mol. Sci. **2014**, 15, 1-x manuscripts; doi:10.3390/ijms150x000x

**OPEN ACCESS**

International Journal of  
Molecular Sciences

ISSN 1422-0067  
www.mdpi.com/journal/ijms

*Type of the Paper (Article, Review, Communication, etc.)*

**Title of the Paper (M\_Title)**

**Firstname Lastname 1,\*, Firstname Lastname 2,†, Firstname Lastname 3,† and   
Firstname Lastname 4**

1 Full Affiliation, Address; E-Mail: author2@email

2 Full Affiliation, Address; E-Mails: author3@email (F.L.); author4@email (F.L.);

**†**These authors contributed equally to this work.

**\*** Author to whom correspondence should be addressed; E-Mail: author1@email;   
Tel.: +1-111-111-111 (ext. 123); Fax: +1-111-111-112.

*Received: / Accepted: / Published:*

**Abstract:** This is the abstract section. One paragraph only (Maximum 200 words).

**Keywords:** keyword; keyword; keyword (3–10 keywords separated by semi colons)

**1. Introduction**

Main text paragraph.

Main text paragraph.

**2. Results and Discussion**

Main text paragraph.

Main text paragraph.

|  |  |
| --- | --- |
| [add an equation here; use MS Word or MathType equation function] | (1) |

(Note: all equations should be completed within a two column table with one line, centered, no boarders, as example see above).

Main text paragraph.

Main text paragraph.

*2.1. This is Subsection Heading*

Main text paragraph.

Main text paragraph.

2.1.1. This is Subsection Heading

Main text paragraph.

Main text paragraph.

**Table 1.** Add a descriptive label of the table here.

[add the table here; use MS Word’s table function]

**Figure 1.** (**a**) Add a descriptive label of the figure here. (**b**) Add a descriptive label of the figure here. (**c**) Add a descriptive label of the figure here.

[add the figure here]

**3. Experimental Section**

3.1. Data

The dataset used in our analysis is due to Debnath *et al* [?] and concerns mutagenic activities of 95 aromatic and heteroaromatic amines in *S. typhimurium* TA98+S9 microsomal preparations. There are 275 descriptors for each compound: among them 97 topostructural, 162 topochemical, 10 3-dimensional and 6 quantum-chemical. Number of revertants per nmol of test culture in log scale (log *R*) is the original response variable. For binary classification we take the 0/1 indicator log *R* being > 0 or < 0 as the response variable in our analysis. At the original scale, this amounts to no. of revertants/nmol being greater/less than 1, respectively.

The results obtained are compared with previous studies on a second, and more diverse, dataset records Ames mutagenicity of 508 chemical compounds (256 mutagens and 252 non-mutagens), and is taken from the CRC Handbook of Identified Carcinogens and Non-carcinogens [?]. The set of descriptors for these compounds includes the above type of descriptors, as well as a large number of Atom-pair (AP) descriptors.

3.2. Methods

3.2.1. Variable selection

Because the number of samples (*n* = 95) is smaller than the number of variables (*p* = 275), regression-based methods of variable selection like stepwise forward and backward selection are not applicable in our scenario. Here we use the Interrelated Two-way Clustering (ITC) algorithm to do the variable selection. This algorithm takes in a number of predictor groups, which are pre-determined or obtained by some known unsupervised clustering method (like K-means). It then hypothesizes that if we include only the important predictors and classify samples independently for each predictor groups, the classification should be identical. Keeping this in mind, an iterative procedure is used to eliminate predictors until the classifications based on different predictor groups achieve a certain level of similarity.

A detailed description of the algorithm can be found in the original paper [Tang et al], and its implementation in QSAR was done by Majumdar *et al* [?].

3.2.2. Hierarchical QSAR and predictive models

We take a hierarchical approach to build the predictive model. Starting from topostructural variables, we keep on including topochemical, 3-dimensional and quantum-chemical variables and check predictive performance of all the models. To tackle high collinearity among different predictors, we use ridge regression to build our predictive models. Given *n* samples and *p* variables, the *n* × *p* data matrix of predictors ***X*** and *n* × *1* vector of 0/1 responses ***Y***, the vector of coefficients obtained by ridge regression is defined as:

***b*** *=* (***X****’****X*** *+ k****I***)-1***Y***

Where *k* is the ridge constant, chosen by cross-validation [?].

While assessing the predictive performance of a model through cross validation, it is essential to not do the variable selection beforehand and the build the model, because that uses information from the test set of compounds in the variable selection step. This results in a synthetic increase of the predictive performance of the model. For this reason, while we did use leave-one-out cross-validation to obtain prediction accuracy, we did both the selection of variables and building ridge regression models for every iteration of the cross-validation procedure.

**4. Conclusions**

Main text paragraph.

Main text paragraph.

**Acknowledgments**

Main text paragraph.

**Author Contributions**

Main text paragraph.

**Conflicts of Interest**

State any potential conflicts of interest here or “The authors declare no conflict of interest”.

**References and Notes**

1. Author 1, A.B.; Author 2, C.D. Title of the cited article. *Journal Title* **2007**, *6*, 100–110.

2. Author 1, A.; Author 2, B. Title of the chapter. In *Book Title*, 2nd ed.; Editor 1, Editor 2, Eds.; Publisher: Publisher Location, Country, 2007; Volume 3, pp. 154–196.

3. Author 1, A.; Author 2, B. *Book Title*, 3rd ed.; Publisher: Publisher Location, Country, 2008;   
pp. 154–196.

4. Author 1, A.B.; Author 2, C. Title of Unpublished Work. Journal Abbreviation, phrase indicating stage of publication.

5. Author 1, A.B.; Author 2, C.D.; Author 3, E.F. Title of Presentation. In *Title of the Collected Work* (if available), Proceedings of the Name of the Conference, Location of Conference, Country, Date of Conference; Editor 1, Editor 2, Eds. (if available); Publisher: City, Country, Year (if available); Abstract Number (optional), Pagination (optional).

6. Author 1, A.B. Title of Thesis. Level of Thesis, Degree-Granting University, Location of University, Date of Completion.

7. Author 1, A.B.; Author 2, C.D. Title of the article. *Abbreviated Journal Name*, Year, Volume, (Page range), doi or other identification number. Available online: http://URL (accessed on Day Month Year).

8. Title of Site. Available online: http://URL (accessed on Day Month Year).

Reference list: We recommend the use of reference management software to prepare the references list (e.g., Endnote, http://www.mdpi.com/files/word-templates/MDPI.ens).

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).